



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(31) International Patent Classification: **A1**
C07C 14/00

(11) International Publication Number: **WO 98/42738**
 (43) International Publication Date: **1 October 1998 (01.10.98)**

(21) International Application Number: **PCT/US98/05311**
 (22) International Filing Date: **19 March 1998 (19.03.98)**

(30) Priority Data:
 60/041,281 21 March 1997 (21.03.97) US
 60/041,276 21 March 1997 (21.03.97) US
 60/042,344 21 March 1997 (21.03.97) US
 60/041,277 21 March 1997 (21.03.97) US
 60/048,355 30 May 1997 (30.05.97) US
 60/048,096 30 May 1997 (30.05.97) US
 60/048,351 30 May 1997 (30.05.97) US
 60/048,134 30 May 1997 (30.05.97) US
 60/048,160 30 May 1997 (30.05.97) US
 60/048,069 30 May 1997 (30.05.97) US
 60/048,131 30 May 1997 (30.05.97) US
 60/048,186 30 May 1997 (30.05.97) US
 60/048,095 30 May 1997 (30.05.97) US
 60/048,187 30 May 1997 (30.05.97) US
 60/048,099 30 May 1997 (30.05.97) US
 60/050,937 30 May 1997 (30.05.97) US
 60/048,352 30 May 1997 (30.05.97) US
 60/048,135 30 May 1997 (30.05.97) US
 60/048,188 30 May 1997 (30.05.97) US
 60/048,094 30 May 1997 (30.05.97) US
 60/048,350 30 May 1997 (30.05.97) US
 60/054,804 5 August 1997 (05.08.97) US

(71) Applicant (for all designated States except US): **HUMAN GENOME SCIENCES, INC.** (US/US); 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors, and
 (75) Inventors/Applicants (for US only): **YOUNG, Paul** (US/US); 122 Beckwith Street, Gaithersburg, MD 20878 (US);
GREENE, John, M. (US/US); 872 Diamond Avenue, Gaithersburg, MD 20878 (US);
FERRIER, Ann, M. (US/US); 13203 L Astoria Hill Court, Germantown, MD 20874 (US);
RUBEN, Steven, M. (US/US); 18528 Hedgely Hills Drive, Olney, MD 20832 (US);
ROSEN, Craig, A. (US/US); 22400 Rolling Hill Road, Laytonville, MD 20882 (US);
DUAN, Roxanne (US/US); 4541 Paintfield Drive, Bethesda, MD 20814 (US);
HU, Jing-Shan (CN/US); 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US);
FLORENCE, Kimberly, A. (US/US); 12805 Atlantic Avenue, Rockville, MD 20851 (US);
OLSEN, Henrik, S. (DK/US); 182 Kendrick Place #24, Gaithersburg, MD 20878 (US);
BENNER, Reinhard (DE/US); 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US);
BREWER, Laurie, A. (US/US); 14920 M. Nebo Road, Poolesville, MD 20837 (US);
MOORE, Paul, A. (GB/US); Apartment 104, 1908 Holly Ridge Drive, McLean, VA 22102 (US);
SHI, Yanguan (CN/US); 437 West Side Drive, Gaithersburg, MD 20878 (US);
LAFLIBRE, David, W. (US/US); 1615 O Street, N.W., #807, Washington, DC 20009 (US);
NI, Jim (CN/US); 5502 Manorfield Road, Rockville, MD 20853 (US).

(74) Agent: **BROOKES, Anders, A. et al.**; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US).

(81) Designated States: **AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GR, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, AR, JP** patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), European patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BI, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
 With international search report.

(54) Title: **87 HUMAN SECRETED PROTEINS**

(57) Abstract

The present invention relates to 87 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AN	Austria	FR	France	LU	Luxembourg	SN	Senegal
AT	Austria	GA	Gabon	LV	Latvia	SE	Sweden
AU	Australia	GB	United Kingdom	MC	Monaco	SG	Singapore
AZ	Azerbaijan	GR	Greece	MD	Moldova	TD	Chad
BA	Bosnia and Herzegovina	HR	Croatia	MG	Madagascar	TG	Togo
BB	Barbados	HN	Honduras	MK	Macedonia	TM	Tajikistan
BE	Belgium	IE	Ireland	ML	Mali	TR	Turkey
BG	Bulgaria	IS	Iceland	MN	Mongolia	TT	Trinidad and Tobago
BI	Burkina Faso	IT	Italy	MR	Mauritania	UA	Ukraine
BJ	Benin	IL	Israel	MV	Maldives	UG	Uganda
BR	Brazil	LV	Latvia	MX	Mexico	UZ	Uzbekistan
BS	Bahamas	LY	Libya	MY	Malaysia	VE	Venezuela
BT	Bhutan	MA	Morocco	NB	Niger	VN	Viet Nam
BY	Belarus	MT	Malta	NE	Nigeria	YU	Yugoslavia
CA	Canada	NP	Nepal	NL	Netherlands	ZW	Zimbabwe
CF	Central African Republic	KR	South Korea	NZ	New Zealand		
CG	Congo	KG	Kyrgyzstan	NO	Norway		
CH	Switzerland	KE	Kenya	PL	Poland		
CI	Cote d'Ivoire	KM	Comoros	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LA	Laos	RU	Russia		
CV	Cape Verde	LC	St. Lucia	SD	Sudan		
CZ	Czech Republic	LI	Liechtenstein	SE	Sweden		
DE	Germany	LK	Sri Lanka	SG	Singapore		
DK	Denmark	LR	Liberia				
EE	Estonia						

87 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO. X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO. X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polynucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods.

Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotyrosine, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, **PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES**, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); **POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS**, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., *Meth Enzymol* 182:626-646 (1990); Rattan et al., *Ann NY Acad Sci* 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table I.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

25 Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

The translation product of this gene shares sequence homology with nucleolin, which is thought to be important in macromolecule binding, as well as some membrane proteins. Preferred polypeptide fragments comprise the amino acid sequence:

30 DPEAADS₁GEPQNKRT₂PDLP₃EEYV₄KEEIQ₅NEEA₆VKKML₇VEAT₈REFEEV₉VVDES
(SEQ ID NO:239); QKLK₁RAEED₂PEAADS₃GEPQNKRT₄PDLP₅EEYV₆KEEIQ₇NEE
AVKKML₈VEAT₉REFEEV₁₀VVDES (SEQ ID NO:240); KAMEKSSL₁TQHSW₂QSLKDR
YLKHL₃RQEHK₄YLLGDAP₅VPSS₆QKLK₇RAEED₈PEAADS₉GEPQNKRT₁₀PDLP₁₁EE
EYV₁₂KEEIQ₁₃NEEA₁₄VKKML₁₅VEAT₁₆REFEEV₁₇VVDES₁₈PPDFEIH₁₉I (SEQ ID NO:241).

Also preferred are the polynucleotide fragments encoding these polypeptide fragments.

This gene maps to chromosome 16, and therefore can be used as a marker in linkage analysis for chromosome 16.

This gene is expressed primarily in brain and kidney and to a lesser extent in wide range of tissues.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cell-cell interaction or cell-matrix interaction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and kidney, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:125 as residues: Met-1 to Trp-10.

15 The tissue distribution and homology to nucleolin indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases involving cell-cell interaction or cell-extracellular matrix interaction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

25 The translation product of this gene shares sequence homology with a porcine zona pellucida protein ZPDS.1711. (See Accession No. R39356.) These two proteins have weak homology with *Drosophila* commissureless and metal homeostasis proteins which are thought to be important in controlling growth cone guidance across the CNS midline and protecting cells against reactive oxygen toxicity, thus, based on homology, it is likely that this gene also be involved in development. Preferred polypeptide

30 fragments comprise the amino acid sequence: LPSYDEAERTKAEAT₁PLV₂GRDEDFVGRDD₃FDAD₄QLRIGND₅GIFML₆TFMAFL₇FNWIG₈FFLS₉FC₁₀LT₁₁TSAA₁₂GRYGAISG₁₃FGLSLIK₁₄WLIV₁₅RFSTY₁₆PGYFDG₁₇QYWL₁₈WVFL₁₉VLG₂₀FL₂₁FL₂₂RGFIN₂₃YAK₂₄VRKM₂₅PETFSN₂₆LR₂₇TVL₂₈FI (SEQ ID NO:242); and/or AGRYGAISG₁FGLSLIK₂WLIV₃RFSTY₄PGYFDG₅QYWL₆WVFL₇VLG₈FL₉FL₁₀RGFIN₁₁YAK₁₂VRKM₁₃(SEQ ID NO:243). Also preferred are polynucleotide fragments encoding these

35 polypeptide fragments. This gene maps to chromosome 5, and therefore can be used in linkage analysis as a marker for chromosome 5.

This gene is expressed primarily in kidney, adrenal gland, brain and to a lesser extent in wide range of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, fertilization control or tissue damages by metabolites or other toxic agents. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and urosecretion system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., kidney, adrenal gland, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to zona pellucida protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for fertility control such as contraceptive development. The homology with metal homeostasis and commissureless genes indicates the gene's function in spermatozoa guidance and protection. It would also be useful for the treatment/diagnosis of tissue damages caused by toxic metabolites and other agents since the gene product is also expressed in urosecretive tissues.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in liver and to a lesser extent in placenta. Preferred polypeptide fragments comprise the amino acid sequence: MKHLSAWNFT KLTELQLWEI FEGS VENCQT L TSYSKL QKYTFSGSTFYI (SEQ ID NO:244). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, digestive and nutrient transport/utilization disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and

circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., liver, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in liver and placenta indicates that the protein product is either an extracellular enzyme or a molecule carrier. Therefore, polynucleotides and polypeptides corresponding to this gene are useful for diagnosis/treatment of digestive and nutrient transport/utilization disorders, including malabsorption and malnutrition.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene shares homology with the *sap47* gene of *Drosophila melanogaster*, a gene which codes for a conserved neuronal protein associated with synaptic terminals. (See Mol. Brain Res. 32:45-54 (1995); see also, Accession No. 929571.) Thus, based on homology, the gene of the present invention also should be associated with synaptic terminals. Preferred polypeptide fragments comprise the amino acid sequence:

FSSDERTSPWESRRVESKATSAACGLWGSGPRRRPASGMFRGLSSWLGLQPP
VAGGGQPNGDAPPEQSETVAESAEEELQAGDELHQAQKDFGNYLFFNFASA
ATKKTIESVAETAQTIKKSVEEGKIDGHDKTIGDFQKEQKKFVEEQHTKKSEA
AVPPWVDTNDEETIQQLALSAADKRNFLRDPAGVQFNFDQMPVALVNL
(SEQ ID NO:245); MRFALVKLVKEEFVW/RNYYFVSLIKQSAQLTALAAQQQA
AGKGGEEQ (SEQ ID NO:246); STSPGVSEFVSDAFDACNLNQEDLRKEMEQL
VLDKKQEETA VLBEDSADWBEKLQQLQEYEVVTESEKRDENWDK (SEQ ID
NO:247); SPWESRRVESKATSAACGLWGSGPRRRPASGMFRGLSSWLGLQQ
PVAGGGQPNGDAPPEQPS (SEQ ID NO:248); PVAGGGQPNGDAPPEQPSSETV
ESAEEELQAGDELHQAQKDFGNYLFFNFASAATKKTIESVAE (SEQ ID NO:
249); and/or FQKEQKKFVEEQHTKKSEAAVPPWVDTNDEETIQQLALSAADR
NFLRDPAGVQFNFDQMPVALVNL (SEQ ID NO:250). Also preferred are
polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in kidney pyramids and to a lesser extent in lung and other tissues of various types. This gene fluxes calcium in human aortic smooth muscle cells, and therefore is involved in signal transduction.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, renal and nervous disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney and/or nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., kidney, lung, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in kidney and lung and homology with sap47 indicates that the protein product has regulatory or direct functions in molecular exchange with body fluids and nervous system signaling. Polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders in kidney and nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 5

The translation product of this gene shares sequence homology with the mouse Ly-9.2 antigen which is thought to be an important cell surface marker in lymphoids, myeloids and hematopoietic progenitors. (See Accession No. gill198932.) Preferred polypeptide fragments comprise the amino acid sequence: PFICVARNPVSRNFSSPI LARKLCEGAA (SEQ ID NO:251); and/or KEDPANTVYSTVEIPKMKMENPHSLLT MPDTPRL (SEQ ID NO:252). Also preferred are polynucleotide fragments encoding these polypeptide fragments. Based on homology, it is likely that this gene is also a cell surface marker, involved in hematopoiesis.

This gene is expressed primarily in activated macrophages, monocytes and T-cells and to a lesser extent in spleen and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., blood cells, and bone marrow, and cancerous and wounded

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:129 as residues: Lys-26 to Tyr-33, Arg-44 to Ile-49, Ser-53 to Lys-71, Lys-86 to Pro-91.

The tissue distribution and homology to Ly-9.2 surface immunoglobulin family indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of immune and hematopoietic disorders. Polypeptides and polynucleotides corresponding to this gene are also be used as a marker for leukemia or a modulator of the functions of the cells of macrophage/monocyte or T-cell types.

FEATURES OF PROTEIN ENCODED BY GENE NO: 6

The translation product of this gene shares sequence homology with the *Drosophila* glutactin gene which is thought to be important in cell-cell interaction or cell-extracellular matrix contact.

This gene is expressed primarily in colon tissue, aorta endothelial cells and to a lesser extent in skin, breast tissue and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of these tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the gastrointestinal tract, vascular system or T-cell development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, cardiovascular system, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., colon, cardiovascular tissue, skin, mammary tissue, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to glutactin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the development and maintenance of the integrity of the basal membrane in the gastrointestinal tract and

11

cardiovascular system. The expression in T-cells also indicate the protein may be involved in T-cell adhesion, cell-cell interaction and development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

5 The translation product of this gene shares sequence homology with MURF4 protein, an ATPase homolog, which is thought to be important in ATP hydrolysis.

This gene is expressed primarily in breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer and non-neoplastic breast diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to MURF4 gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neoplastic or non-neoplastic breast diseases because ATPase like protein may be involved in changed metabolic states of the breast.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 8

This gene shares homology to the alcohol dehydrogenase gene. Preferred polypeptide fragments comprise the amino acid sequence: ASA VLLDLPNSG GEQAQKLGNNCFAPADVTSEKDVQTAALAKGKGRVDVAVNCAGIAVAS KTYNLKKGQTHLTEDFORVLDDVNLMTGNVIRLVAGEMGQNEPDQGGQGVINTASVAAFEGQVGQAAYSASKGGIVGMTLPARDLAPIGIRVMTIAPQLFQGPLTSLPEKVCNFLASQVPPSRLGDPAEYAHLVQALINPPLNGEVIRLDGARMQ P (SEQ ID NO:253); and/or SVAAFEGQVGQAAYSASKGGIVGMTPLA (SEQ ID NO:254). Polynucleotides encoding these fragments are also encompassed by the invention. Other groups have also recently cloned this gene, recognizing its homology to alcohol dehydrogenase. (See Accession No. 1778355.) Moreover, a second group

12

recently cloned the mouse homologue of this gene. (See Accession No. 2078284.) They found that the mouse homologue binds to amyloid beta-peptide and mediates neurotoxicity in Alzheimer's disease, calling the protein ERAB. This gene maps to chromosome X, and therefore can be used in linkage analysis as a marker for chromosome X. Therefore, mutations in the translated product of this gene may be involved in Alzheimer's disease in humans, as well as other sex linked diseases. This gene can be used as a diagnostic marker for these diseases.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:132 as residues: Arg-45 to Ser-53.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 9

The translation product of this gene shares weak sequence homology with rat N-methyl-D-aspartate receptor subunit and other proline-rich proteins which are thought to be important in neurotransmission or protein-protein interaction.

15 This gene is expressed primarily in synovial hypoxia and to a lesser extent in ovary, senescent cells and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, synovial hypoxia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the synovia and brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., synovial tissue, ovary and other reproductive tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in synovial hypoxia and nerve tissues, and homology to N-methyl-D-aspartate receptor subunit and other proline-rich proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of synovial hypoxia and other synovial disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

This gene is expressed primarily in prostate and to a lesser extent in placenta and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male and female infertility, cancer, and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system and neoplasia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., prostate, placenta, ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:134 as residues: Pro-17 to Met-23, Ala-30 to Trp-38, Ile-49 to Trp-54, Lys-68 to Gly-74, Thr-93 to Gly-99, Met-126 to Glu-132, Gly-173 to Ser-178, Lys-205 to Tyr-214.

The tissue distribution of this gene in the prostate, placenta and ovary indicates that this gene product is useful for treatment/diagnosis of male or female infertility, endocrine disorders, fetal deficiencies, ovarian failure, amenorrhea, ovarian cancer, benign prostate hyperplasia, prostate cancer, and other forms of cancer of the reproductive system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in the thyroid and to a lesser extent in the pineal gland. This gene maps to chromosome 10, and therefore can be used as a marker in linkage analysis for chromosome 10. Preferred polypeptide fragments comprise the amino acid sequence: HPIEWADNAATLSQFY (SEQ ID NO:256); CWIKYCLTLMQN AQLSMQDNIG (SEQ ID NO:257); KVSYLRLDFEEARELFLGQHYYVF (SEQ ID NO:258); MERRCKMHKRXIAMLEPLTYDLNPQ (SEQ ID NO:259); and/or SHIV KKINLNKSALKY YQLFLD (SEQ ID NO:260). Also preferred are polynucleotides encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, thyroid and pineal gland disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., thyroid and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:135 as residues: Ser-2 to Ser-8, Thr-38 to Arg-44.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/detecting immune disorders such as arthritis, asthma, immune deficiency diseases (e.g., AIDS), and leukemia, as well as treating/detecting thymus disorders (e.g., Graves Disease, lymphocytic thyroiditis, hyperthyroidism, and hypothyroidism), and treating/detecting pineal gland disorders (e.g., circadian rhythm disturbances associated with shift work, jet lag, blindness, insomnia and old age).

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed primarily in lung and tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pulmonary or immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., pulmonary tissue, and tonsils, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily

15

fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:136 as residues: Gln-28 to Gly-49.

The tissue distribution of this gene only in lung indicates that it could play a role in the treatment/detection of lung lymphoma or sarcoma formation, pulmonary edema and embolism, bronchitis and cystic fibrosis. Its expression in tonsils indicates a potential role in the treatment/detection of immune disorders such as arthritis, asthma, immune deficiency diseases (e.g., AIDS), and leukemia, in addition to the treatment/detection of tonsillitis.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed primarily in lymphoid, myeloid and erythroid cells. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, myeloid cells, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The predominant tissue distribution of this gene in hematopoietic cell types indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, asthma, immunodeficiency diseases and leukemia. Preferred embodiments of the present invention are polypeptide fragments comprising the amino acid sequence: FTHLSTCLLSLLVRMSGELLARASPSI
30 CALDSSCFVEYCSSYSSCFHLHQHPSLLDHLCO (SEQ ID NO:261); or FLLL
ARASPSICALDSSCFVQEY (SEQ ID NO:262). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene is homologous to the *Drosophila Regena* (*Rga*) gene. (See Accession No. 1658504.) This *Drosophila* gene is thought to be a homolog of the global negative

16

transcriptional regulator NOT2 (CDC36) from yeast, which modifies gene expression and suppresses position effect variegation. Preferred polypeptide fragments comprise the amino acid sequence: PDGRVTNIPQGMVTDQFGMIGLLTFRAETDPGMVHL
5 ALGSDLTTLGLNLNS (SEQ ID NO:263); VHLALGSDLTTLGLNLNSPENLYP
(SEQ ID NO:265); EDLLFPLYMNGDYLQLLAVELFNRDWRVHKEERVWI
TR (SEQ ID NO:264); and/or HNEDFPALPOS (SEQ ID NO:266).

This gene is expressed primarily in placenta and to a lesser extent in infant brain.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurological system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:138 as residues: Leu-9 to Tyr-15, Asp-34 to Gln-46, Pro-51 to Asp-57, Gly-88 to Thr-104, Thr-123 to Ser-128.

25 The tissue distribution of this gene indicates that it could be used in the detection and/or treatment of neurological disorders such as such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, and panic disorder.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 15

This gene is expressed primarily in adrenal gland tumor and osteoclastoma. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, endocrine and bone disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

5 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system and in bone, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adrenal gland, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:139 as residues: Ile-52 to Trp-57.

10 The tissue distribution of this gene indicates that it may be involved in the treatment and/or detection of adrenal gland tumors, osteosarcomas, endocrine disorders and bone disorders.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 16

The translation product of this gene shares sequence homology with the FK506 binding protein, a protein which plays an important role in immunosuppression. (See Accession No. M75099.) Specifically, a 12-kDa FK506-binding protein (FKBP-12) is a cytosolic receptor for the immunosuppressants FK506 and rapamycin. (See, Proc. Natl. Acad. Sci. 88: 6677-6681 (1991).) Thus, based on homology, it is likely that this gene also has immunosuppression activity. Preferred polypeptides comprise the amino acid sequence: GRIDTSLTRDPLVIELGQKQVIPGLEQSLDMCVGEKRAIPSH LAYGKRGPPSPADAVVQYDVIELALIR (SEQ ID NO:267); and/or IHTGSLV DGR IDTS (SEQ ID NO:268). Also preferred are the polynucleotide fragments encoding these polypeptides.

25 This gene is expressed primarily in melanocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:140 as residues: Ala-118 to Phe-124, Arg-178 to Lys-201.

5 The tissue distribution and homology to the FK506 binding proteins which are believed to a role in immunosuppression mediated by the immunosuppressant drugs rapamycin and cyclosporin, indicates that this gene could serve as a novel target for the identification of novel immunosuppressant drugs.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The translation product of this gene shares sequence homology with the rat calcium-activated potassium channel rSK3, which is thought to be important in regulating vascular tone. (See Accession No. gi2564072, gi1575663, and gi1575661.) Although homologous to these proteins, this gene contains an 18 amino acid insert, not previously identified in the homologs. Preferred polypeptide fragments comprise the amino acid sequence: CESPSPAQPSGSSLPAYYH (SEQ ID NO:269). Also preferred are the polynucleotide fragments encoding these polypeptides.

This gene is expressed primarily in B-cells, frontal cortex and endothelial cells. Therefore, polynucleotides and polypeptides of the invention are useful as

20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular (hyper/hypotension, asthma, pulmonary edema, pneumonia, heart disease, restenosis, atherosclerosis, stroke, angina and thrombosis) and neurological disorders. Similarly, polypeptides and antibodies directed to these

25 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, brain and other tissue of the nervous system, and endothelium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine,

30 synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID

35 NO:141 as residues: Glu-72 to Gly-82, His-90 to Val-95, Gln-168 to Lys-174, Val-202 to Ser-212.

19

The tissue distribution and homology to calcium-activated potassium channels indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of vascular disorders (hyper/hypotension, atherosclerosis, stroke, angina, pneumonia, heart disease, restenosis, atherosclerosis, stroke, angina and thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

This gene is expressed primarily in smooth muscle and to a lesser extent in brain (amygdala, corpus colosum, hippocampus).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular (hypertension, heart disease, atherosclerosis, pulmonary edema, restenosis, atherosclerosis, stroke, angina, thrombosis, and wound healing), and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and neurological systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., smooth muscle, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 142 as residues: Lys-43 to Arg-49, Tyr-58 to Glu-65.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cardiovascular disorders (hypertension, heart disease, atherosclerosis, pulmonary edema, restenosis, atherosclerosis, stroke, angina, thrombosis, and wound healing). Expression in brain indicates a role in the treatment and diagnosis of behavioral or neurological disorders, such as depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 19

This gene is expressed primarily in T-cells (Jurkats, resting, activated, and

20

anergic T-cells), endothelial cells, pineal gland, and to a lesser extent in a variety of other tissues and cell types. Preferred polypeptide fragments comprise the amino acid sequence: BEAGAGRRCSHGARPAGLQNEGLGLGDDPDHTDTSRKSQRINN WKESKHKVMASASARGNQDKDAHFPSPKSLFCPEKSKLHHRAEISK (SEQ ID NO:270); and/or SKQRINNVKESKHKVMASASAR (SEQ ID NO:271). Also preferred are the polynucleotide fragments encoding these polypeptides.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, immune and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, neurological and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, endothelial cells, and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:143 as residues: Phe-71 to Arg-76, Pro-82 to His-87, Glu-103 to Ala-111.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. In addition, expression in the pineal gland might suggest a role in the diagnosis of specific brain tumors and treatment of neurological disorders. Endothelial cell expression might suggest a role in cardiovascular or respiratory/pulmonary disorders or infections (atherosclerosis, pneumonia).

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene is expressed primarily in brain and embryo and to a lesser extent in leukocytes. This gene maps to chromosome 15, and therefore can be used as a marker in linkage analysis to chromosome 15.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g. cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:144 as residues: Met-1 to Gly-8.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. The expression in the brain -- and in particular the fetal brain -- would suggest a possible role in the treatment and diagnosis of developmental and neurodegenerative diseases of the brain and nervous system (depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior).

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

This gene is expressed primarily in brain, kidney, lung, liver, spleen, and a variety of leukocytes (especially T-cells) and to a lesser extent in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemias, lymphomas, autoimmune, immunosuppressive, and immunodeficiencies, hematopoietic disorders, as well as renal disorders, and neoplasms. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal, pulmonary, immune, and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,

brain and other tissue of the nervous system, kidney, pulmonary tissue, liver, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of renal conditions, such as acute renal failure, kidney fibrosis, and kidney tubule regeneration. The expression in leukocytes and other immune tissues indicates a role in immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. The expression in the brain -- and in particular the fetal brain -- indicates a possible role in the treatment and diagnosis of developmental and neurodegenerative diseases of the brain and nervous system (depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior).

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

This gene is expressed primarily in skin (fetal epithelium, keratinocytes and skin). This gene also maps to chromosome 19, and therefore can be used in linkage analysis as a marker for chromosome 19.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, skin cancers (e.g., melanomas), eczema, psoriasis or other disorders of the skin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., skin and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:146 as residues: Pro-28 to Glu-35. Ser-39 to Phe-44, Ala-94 to Gln-99.

23

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of skin cancers (e.g., melanomas), eczema, psoriasis or other disorders of the skin.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 23

This gene maps to chromosome 11. Another group recently isolated this same gene, associating the sequence to the region thought to harbor the gene involved in Multiple Endocrine Neoplasia Type 1, or MEN 1. (See Accession No. 2529721 and Genome Res. 7(7), 725-735 (1997), incorporated herein by reference in its entirety.) Preferred polypeptide fragments comprise the amino acid sequence: LFWACLNERA AQLPNTAXAGYQCPSGNGPS (SEQ ID NO:272).

This gene is expressed primarily in epididymus, pineal gland, T-cells, as well as fetal epithelium, lung and kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, metabolic mediated disorders, and MEN. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, renal, neurological and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., epididymus and other reproductive tissue, pineal gland, T-cells and other blood cells, epithelium, lung, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of developmental deficiencies or abnormalities as well as a host of different disorders which arise as a result of conditions in the indicated tissues or cell types. An area of particular interest is in the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. The expression in the brain, and in particular the fetal brain, would suggest a possible role in the treatment and diagnosis of

24

developmental and neurodegenerative diseases of the brain and nervous system (depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior). Respiratory/pulmonary disorders, such as asthma, pulmonary edema are also potential therapeutic areas, as well as renal conditions such as acute renal failure, kidney fibrosis and kidney tubule regeneration. Moreover, this gene can be used in the treatment and/or detection of MEN 1.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 24

This gene is expressed primarily in fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemia, lymphoma, AIDS, hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., spleen and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 25

A closely related homolog of this gene was recently cloned by another group, calling the gene CDO, an oncogene-, serum-, and anchorage-regulated member of the Ig/fibronectin type III repeat family. (See Accession No. 2406628, and J. Cell Biol. 138(1): 203-213 (1997), herein incorporated by reference in its entirety.) Preferred polypeptide fragments comprise the amino acid sequence: FYIYYRPTDSDNDSDYKK DMVEGDKYVHSISHLQPETSYDIKMQCNEGEEFSNVVICETKARKSSGQP GRLPPTLAPQPPLPETIERPVGTGAMVARSDDLPLYLVGVVLSIVLIVITFPE CLWRWSKQKHITDLAGFRSALPPSCPVTMVPGLGPHQA VDSPTSVASVD

GPVLM (SEQ ID NO:273); or YYYRPTSDNDSDYKDKMVEGDKYWHSHSLQ
PETSYDKMQCFNEGSEFSNVMICETKARKS (SEQ ID NO:274).

This gene is expressed primarily in fetal lung and kidney, human embryo and osteoclastoma stromal cells and to a lesser extent in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental disorders and cancers, as well as pulmonary and renal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the respiratory/pulmonary, skeletal and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung, kidney, embryonic tissue, and bone cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:149 as residues: Thr-5 to Pro-18, Ala-76 to Thr-84.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of: osteoporosis, fracture, osteosarcoma, ossification, and osteonecrosis, as well as respiratory/pulmonary disorders, such as athesma, pulmonary edema, and renal conditions such as acute renal failure, kidney fibrosis and kidney tubule regeneration.

FEATURES OF PROTEIN ENCODED BY GENE NO: 26

This gene is homologous to the HIV envelope glycoprotein. (See Accession No. 2641463.) Preferred polypeptide fragments comprise the amino acid sequence: NVRALLHRMPEPPKINTAKFNNNKKNLSL (SEQ ID NO:275).

This gene is expressed primarily in pineal gland and skin, and to a lesser extent in lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, neurological and behavior disorders; respiratory/pulmonary disorders, such as athesma, pulmonary edema; skin conditions such as eczema, psoriasis, acne and skin cancer, as well as AIDS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and respiratory systems, as well as skin and AIDS, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells; pineal gland, epidermis, and pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:150 as residues: Gln-15 to Gln-20.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions which affect the above tissues, such as: skin cancer, eczema, psoriasis, acne, alhesma, pulmonary edema, neuro-degenerative or developmental disorders such as Alzheimer's, depression, schizophrenia, dementia, and AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

Preferred polypeptide encoded by this gene comprise the following amino acid sequence: NTNQREALQYAKNFQPFALNHQKDIQVLMGSLVYLRQGIENSPYVHL LDANQWADICDIFTRDACAALLGLSVESPLSVSFSAGCVALPALNIKA VIEQRQC TGVWNQKDELPIEVDLGKKCWYHSIFACPILRQQTTDNNPPMKLVCGHISRDLNKMFGSKLKCPYCPMEQSPGDAKQIFF (SEQ ID NO:276). Polynucleotides encoding such polypeptides are also provided as are complementary polynucleotides thereto.

This gene is expressed primarily in liver (adult and fetal) and spleen tissue, and to a lesser extent in placenta, T helper cells, kidney tumor, ovarian tumor, melanocytes and fetal heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and developmental diseases and disorders and liver diseases such as liver cancer. Similarly, polypeptides and antibodies directed to these

27

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, circulatory and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, placenta, blood cells, kidney, ovary and other reproductive tissue, melanocytes, and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of growth, hematopoietic and immune system disorders particularly related to the liver.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 28

The translation product of this gene shares sequence homology with prostaglandin transporter which is thought to be important in metabolic and endocrine disorders. See, for example, Gastroenterology Oct;109(4):1274-1282 (1995). Preferred polypeptides encoded by this gene comprise the following amino acid sequence:
SYLSACFAGCNSNTNLTGACCLTTYPAENATVVPKCPSPQCQAEFLTFLCYMCI
CSLIGAMARHP (SEQ ID NO:277); and/or PSVILLRTVSPELKSVALGVLFLRL
LGFTPLPLFGAGIDSTCLFWSTFCGEQACVLYDNVVYRRLVYSIALAKSFAFI
(SEQ ID NO:278).

This gene is expressed primarily in hematopoietic and brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic, immune and endocrine diseases and disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic, immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., endocrine tissue, hematopoietic tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

28

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to prostaglandin (and anion) transporter indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of endocrine, metabolic, immune and kidney disorders.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 29

This gene is expressed primarily in early stage human lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, growth and respiratory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developmental and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:153 as residues: Val-50 to Trp-55.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of respiratory and growth diseases and disorders.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 30

The translation product of this gene shares sequence homology with human

DNA helicase which is thought to be important in accurate and complete DNA replication in creation of new cells. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: QSLFTREVRVGVPVVDLDAQGRARA
SLCXXYNWRYKNLGNLPHYQLPESTANAGLLYDFQLINVEDEQVGGESEPN
PYFYQNLGEAEYVVALFMYMCLLGYPADKLSLLTYNGQKHLRDIINRRCONN
PLGRPNKVTIVDRFQGGQNDYLLSLVTRAVGHLRDVRLVVAAMSRAR
(SEQ ID NO:279); and/or LVKEAKIAMTCTHAALKRHDLVKLGRKYDNLMBE
AAQLIEITFIPLLQNPDGFSRLKRWIMGDHHLPPVI (SEQ ID NO:280).

This gene is expressed primarily in testes tumor and to a lesser extent in adrenal

29

gland tumor and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers and endocrine/growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine, developmental, and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., testes and other reproductive tissue, adrenal gland, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to DNA helicasase indicates that the protein products of this gene are useful for study, treatment, and diagnosis of many cancer types, including testicular cancer, as well as disorders involving endocrine function and normal growth and development.

20

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

The translation product of this gene shares sequence homology with BID-apoptotic death gene (mouse), Genbank accession no. PID g1669514, which is thought to be important in programmed cell death.

This gene is expressed primarily in jurkat membrane bound polysomes and activated neutrophils and to a lesser extent in endothelial cells and human cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers and other proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, endothelium, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

30

urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:155 as residues: Glu-4 to Leu-11, Cys-28 to Arg-35, Gln-50 to His-66, Glu-73 to Gln-79, Gly-94 to Ser-100, Arg-114 to Asp-126, Pro-139 to Lys-146.

The tissue distribution and homology to BID-apoptotic death gene indicates that the protein products of this gene are useful for study of cell death, and treatment and diagnosis of proliferative disorders and cancers. Apoptosis - programmed cell death - is a physiological mechanism involved in the deletion of peripheral T lymphocytes of the immune system, and its dysregulation can lead to a number of different pathogenic processes. Diseases associated with increased cell survival, or the inhibition of apoptosis, include cancers (such as follicular lymphomas, carcinomas with p53 mutations, and hormone-dependent tumors, such as breast cancer, prostate cancer, Kaposi's sarcoma and ovarian cancer); autoimmune disorders (such as systemic lupus erythematosus and immune-related glomerulonephritis rheumatoid arthritis) and viral infections (such as herpes viruses, pox viruses and adenoviruses), inflammation; graft vs. host disease, acute graft rejection, and chronic graft rejection. Diseases associated with increased apoptosis include AIDS; neurodegenerative disorders (such as Alzheimer's disease, Parkinson's disease, Amyotrophic lateral sclerosis, Retinitis pigmentosa, Cerebellar degeneration); myelodysplastic syndromes (such as aplastic anemia), ischemic injury (such as that caused by myocardial infarction, stroke and reperfusion injury), toxin-induced liver disease (such as that caused by alcohol), septic shock, cachexia and anorexia. Thus, the invention provides a method of enhancing apoptosis in an individual by treating the individual with a polypeptide encoded by this gene.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 32

The translation product of this gene shares sequence homology with human fructose transporter which is thought to be important in normal metabolic function and activity.

This gene is expressed primarily in T-cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

35

not limited to, leukemia and other cancers, and metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic, lymph and metabolic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:156 as residues: Pro-22 to Gly-48, Ser-54 to Pro-61.

The tissue distribution indicates that the protein products of this gene are useful for study of mechanisms leading to cancer, treatment and diagnosis of cancerous and pre-cancerous conditions; as well as the study and treatment of various metabolic diseases and disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is expressed primarily in human meninges.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and other disorders of the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., meninges and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:157 as residues: Asn-23 to Pro-31.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of disorders of the CNS and inflammatory responses.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene is expressed primarily in activated monocytes and wound healing tissues and to a lesser extent in fetal epithelium.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and inflammatory disorders and wound healing and tissue repair dysfunctions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, epithelial and gastrointestinal systems, and healing wounds, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., monocytes and other blood cells, and epithelium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:158 as residues: Ala-28 to Ala-33, Gly-35 to Gln-45.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis, study and treatment of immune and inflammatory disorders and wound healing dysfunctions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

This gene is expressed primarily in human osteosarcoma and prostate cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, skeletal and neoplastic conditions such as bone and prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., bone, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial

fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:159 as residues:

- 5 Ser-14 to Gly-22, Leu-37 to Gln-43.
The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of skeletal disorders and cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

- 10 This gene encodes a protein which is highly homologous to a protein called congenital heart disease protein 5, presumably implicated in congenital heart disease (see Genbank PID g2810996).

- 15 This gene is expressed primarily in Hodgkin's lymphoma, erythroleukemia cells, and TNF activated synovial fibroblasts, to a lesser extent in ovarian cancer, cerebellum, spleen, fetal liver and placenta and finally to a lesser extent in various other mesenchymal tissues.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer, immune, hematopoietic and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune,

- 25 hematopoietic and cardiovascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., heart and other cardiovascular tissue, lymphoid tissue, blood cells, bone marrow, ovary and other reproductive tissue, brain and other tissue of the nervous system, spleen, liver, and mesenchymal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:160 as residues: Lys-41 to Met-49, Gln-54 to Glu-59, Glu-76 to Thr-88.

- 35 The homology of this gene and translation product to congenital heart disease protein 5 indicates a role for this protein in the diagnosis, prognosis and/or treatment of

- heart disease or other cardiovascular related disorders. In addition, predominant expression in cells associated with the immune and hematopoietic system indicates a role for this protein in the treatment, diagnosis and/or prognosis of immune and autoimmune diseases, such as lupus, transplant rejection, allergic reactions, arthritis, asthma, immunodeficiency diseases, leukemia, AIDS, thymus disorders such as Graves Disease, lymphocytic thyroiditis, hyperthyroidism and hypothyroidism, graft versus host reaction, graft versus host disease, transplant rejection, myelogenous leukemia, bone marrow fibrosis, and myeloproliferative disease. The protein could also be used to enhance or protect proliferation, differentiation and functional activation of hematopoietic progenitor cells such as bone marrow cells, which could be useful for cancer patients undergoing chemotherapy or patients undergoing bone marrow transplantation. The protein may also be useful to increase the proliferation of peripheral blood leukocytes, which could be useful in the combat of a range of hematopoietic disorders including immunodeficiency diseases, leukemia, and septicemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 37

This gene is expressed primarily in ovarian cancer.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, urogenital neoplasias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:161 as residues: Asn-22 to Asn-27.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of ovarian and other tumors.

35

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 38

The translation product of this gene shares sequence homology with zinc finger proteins.

This gene is expressed primarily in various fetal, cancer, and endothelial lines. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of immune and developmental conditions and cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 39

This gene is expressed primarily in fetal, infant, and adult brain and to a lesser extent in other brain and endocrine organs and blastomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain tumors and neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, endocrine tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the

36

disorder.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of brain cancer and other neurological disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

The translation product of this gene shares sequence homology with vesicular glycoproteins and lectins. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: DTYPNEKQQRVFPXXSAMVNNGLSYDHER DGRPTLGGCCXAVRNLYHYDTFLVIRVYKRALTIMMDIDGKHEWRDCIEYPGV RLPRGYFTGTSITGDLSDNHDVISLKLFELETVERTPEEE (SEQ ID NO:281); and/or LKREHLSKPYQGVGTGSSSLWNLGMNAMVMTQYRLTPDMQSKQGA LWNRVCEFLRDWELQVHFKEHQGQKKNLHGDLAIWYT (SEQ ID NO:282).

This gene is expressed primarily in infant brain and to a lesser extent in various normal and transformed neural, endocrine, and immune organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and neurodevelopmental conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hormonal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, endocrine tissue, and tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:164 as residues: Pro-64 to Gly-71, Gly-94 to Leu-100, Thr-110 to Pro-116, Thr-135 to Arg-145, Glu-164 to Glu-171, Asp-204 to Asp-211, Arg-253 to His-261, Asn-312 to Tyr-323.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of mental retardation and other neurological disorders and neoplasias.

FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene displays homology to the glycosyltransferase family, which catalyze the addition of sialic acids to carbohydrate groups which are present on glycoproteins.

This gene is expressed primarily in smooth muscle and to a lesser extent in pineal gland, fetal liver, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, gastrointestinal injury, inflammatory and neurodegenerative conditions.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., smooth muscle, pineal gland, liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:165 as residues: Ser-12 to Trp-21, Arg-24 to Pro-32, Asp-73 to Lys-82, Lys-90 to Ala-97.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of neurodegenerative and growth disorders and gastrointestinal repair.

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

The translation product of this gene shares sequence similarity with metallothionein polypeptides. See, for example, Proc. Natl. Acad. Sci. U S A 1992 Jul 15;89(14):6333-6337. Metallothioneins are believed to inhibit neuronal survival among other biological functions. Based on the sequence similarity (especially the conserved cysteine motifs characteristic of the metallothionein family) the translation product of this gene is expected to share certain biological activities with other members of the metallothionein polypeptide family. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: PGTQQSALHHDPGCANCSSRFCD

CSPPACQC (SEQ ID NO:283).

This gene is expressed exclusively in placenta and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, liver, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to metallothionein indicates that the protein products of this gene are useful for diagnosis and treatment of immune and hematopoietic system disorders and neurological diseases, especially in fetal development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 43

Preferred polypeptides encoded by this gene comprise the following amino acid sequence: FLYDVLMMXHEAVMRTHQQLPDPEFPS (SEQ ID NO:284).

This gene is expressed primarily in T-cells and synovial tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, and T-cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treatment and diagnosis of disorders of the immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 44

The translation product of this gene shares sequence similarity with several methyltransferases (e.g., see Genbank gill1065505).

This gene is expressed primarily in ovary, thymus, infant adrenal gland, tissues of the nervous system and the hematopoietic tissue, and to a lesser extent in adipose tissue and many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the reproductive system, the endocrine system, the hematopoietic system and the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, endocrine, CNS and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., ovary and other reproductive tissue, thymus, adrenal gland, brain and other tissue of the nervous system, hematopoietic tissue, and adipose tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:168 as residues: Ser-3 to Gly-12, Asp-19 to Arg-31, Tyr-70 to Tyr-77, Asn-130 to Lys-140, Pro-165 to Gln-170, Pro-192 to Lys-199, Leu-216 to Glu-227, Glu-254 to Phe-281.

The tissue distribution and homology to methyltransferase indicates that the protein products of this gene are useful for diagnosis and treatment of disorders of the CNS, the hematopoietic system and reproductive organs and tissues. For example, the abundant expression in the ovary may indicate that the gene product can be used as a hormone with either systemic or reproductive functions; as growth factors for germ cell maintenance and in vitro culture; as a fertility control agent; remedy for sexual dysfunction or sex development disorders; diagnostics/treatment for ovarian tumors, such as serous adenocarcinoma, dysgerminoma, embryonal carcinoma,

choriocarcinoma, teratoma, etc). The expression in thymus may indicate its utilities in T-cell development and thus its applications in immune related medical conditions, such as infection, allergy, immune deficiency, tissue/organ transplantation, etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

The translation product of this gene shares sequence homology with cytochrome C oxidase which is thought to be important in metabolic function of cells. This gene has now recently been published as estrogen response gene. See Genbank accession no. AB007618 and Mol. Cell. Biol. 18 (1), 442-449 (1998). See also J Immunol. Mar 1:154(5): 2384-2392 (1995), where the mouse homologue was published and implicated in silicosis.

This gene is expressed primarily in adipose tissue, kidney and fetal brain and to a lesser extent in several other tissues and organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic diseases involving especially adipose tissue, brain and kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS and vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adipose tissue, kidney, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:169 as residues: Thr-5 to Ser-14.

The tissue distribution and homology to cytochrome C oxidase, estrogen response gene product and silicosis related gene product indicates that the protein products of this gene are useful for diagnosis and treatment of metabolic disorders in the CNS, adipose tissue and kidney, particularly silicosis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 46

The translation product of this gene shares sequence homology with reticulocalbin. See, for example, J. Biochem. 117 (5), 1113-1119 (1995). Based on the

sequence similarity, the translation product of this gene is expected to share certain biological activities with reticulocalbin, e.g., Ca++ binding activities. This gene product is sometimes hereinafter referred to as "Reticulocalbin-2".

This gene is expressed primarily in breast, endothelial cells, synovial, heart and smooth muscle cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the breast, vascular and skeletal/cardiac muscular system.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast, vascular and skeletal-muscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, endothelial cells, synovial tissue, heart and other cardiovascular tissue, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:170 as residues: Gly-16 to Arg-32, Ala-42 to Asn-50, Glu-66 to Gln-76, Arg-85 to Gly-94, Thr-108 to Asp-115, Trp-121 to Gly-130, Leu-137 to His-144, Glu-155 to Lys-161, Asp-175 to Ser-180, Glu-209 to Gly-217, Glu-232 to Glu-237, Thr-243 to Asp-261, Glu-287 to Arg-295.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of diseases of the vascular and skeletal/cardiac muscular system. The homology of the gene with reticulocalbin indicates its biological function in regulating calcium store, a particularly important function in muscular cell types. The gene expression in the heart may indicate its utilities in diagnosis and remedy in heart failure, ischemic heart diseases, cardiomyopathy, hypertension, arrhythmia, etc. The abundant expression in the breast may indicate its applications in breast neoplasia and breast cancers, such as fibroadenoma, papillary carcinoma, ductal carcinoma, Paget's disease, medullary carcinoma, mucinous carcinoma, tubular carcinoma, secretory carcinoma and apocrine carcinoma; juvenile hypertrophy and gynecomastia, mastitis and abscess, duct ectasia, fat necrosis and fibrocystic diseases, etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 47

The translation product of this gene shares weak sequence homology with H+-transporting ATP synthase which is thought to be important in cell metabolism or signal transduction.

5 This gene is expressed only in testis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of some types of diseases and conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and hematopoietic tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Since only one out of about a million expressed sequence tag is found in testes indicates that its expression is low and selectively in testes. Since some of the genes only expressed in testes are usually expressed in brain or in certain induced hematopoietic cells/tissues, it is speculated that this gene to be expressed in brain or hematopoietic cells/tissues and is useful for diagnosis and treatment of disorders these systems.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 48

The translation product of this gene shares sequence homology with human polymeric immunoglobulin receptor (accession No.X73079) which is thought to be important in antibody recognition and immune defenses. In one embodiment, polypeptides of the invention comprise the sequence GWYWCG (SEQ ID NO:285). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in placenta and to a lesser extent in corpus callosum and fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

43

not limited to, disorders of the immune system, e.g. autoimmune diseases and immunodeficiency. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:172 as residues: Tyr-37 to Cys-49, Gly-51 to Tyr-56, Lys-88 to Trp-93, Leu-130 to Glu-136.

The tissue distribution and homology to human polynuclear immunoglobulin receptor indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, e.g. autoimmune diseases and immunodeficiencies.

FEATURES OF PROTEIN ENCODED BY GENE NO: 49

20 This gene is expressed in thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., thymus and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, e.g. autoimmunity and immunodeficiency.

44

FEATURES OF PROTEIN ENCODED BY GENE NO: 50

Preferred polypeptide encoded by this gene comprise the following amino acid sequence: MKVGARIRVKMSVKNKAPVSTHWRPWPAEWPQMLHLAQERTE VKSRPLGLAGFIQDSKTRKPLEQETMSAADTALWPYGHGNREHQENELQKY LQYKDMHLDSQSLGHTHTLQGSNLTALNI (SEQ ID NO:286). Polynucleotides encoding this polypeptide are also provided as complementary polynucleotides thereto.

This gene is expressed primarily in adrenal gland, pituitary, T helper cells, and breast cells and to a lesser extent in a wide variety of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of the some diseases and conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adrenal gland, pituitary, T-cells and other blood cells, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:174 as residues: Gln-39 to Ser-47, Arg-57 to Glu-67, Tyr-82 to Gln-95.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of a wide range of disorders, such as immune and endocrine disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 51

The translation product of this gene shares sequence homology with human Sop2p-like protein which is important in cytoskeleton structure. In one embodiment, polypeptides of the invention comprise the sequence SLHKNSVSQISVLSGGKAKCS QFCTGMDGGMSTWDVKSLESALKD LKI (SEQ ID NO:287). Polynucleotides encoding this polypeptide are also encompassed by the invention. This gene maps to chromosome 7. Therefore, polynucleotides of the invention can be used in linkage

analysis as a marker for chromosome 7.

This gene is expressed primarily in immune and hematopoietic tissues/cells and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological and hematopoietic disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., immune and hematopoietic tissue/cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:175 as residues: Lys-49 to Gln-54, Ala-61 to Arg-66, Lys-82 to Lys-87, Glu-126 to Val-133, His-136 to Ile-141, Glu-175 to Ser-187, Asp-286 to Leu-296, Ala-298 to Ser-310.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immunological, hematopoietic, and inflammatory disorders, e.g., immunodeficiency, autoimmunity, inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 52

The translation product of this gene shares sequence homology with *Caenorhabditis elegans* R53.5 gene encoding a putative secreted protein without known function.

This gene is expressed primarily in endothelial cells, brain and several highly vascularized, and tumor tissues and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, aberrant angiogenesis and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular and brain system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, brain and other tissue of the nervous system, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:176 as residues: Thr-43 to Asn-60, Thr-106 to Phe-115, Asp-122 to Arg-133, Arg-186 to Asp-192, Leu-211 to Lys-216.

The tissue distribution and homology to a *C. elegans* secreted protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of disorders in vascular or brain system, e.g. aberrant angiogenesis, ischemia, neurodegeneration, etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

In one embodiment, polypeptides of the invention comprise the sequence EASKSSHAGLDLFSVAACHRF (SEQ ID NO:288). Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed primarily in T-cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lymphocytic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the lymphoid system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:177 as residues: Pro-3 to Thr-8, Arg-37 to Asp-46.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, treatment, and cure of lymphocytic disorders.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 54

The translation product of this gene shares sequence homology with secreted cartilage matrix protein, a major component of the extracellular matrix of nonarticular cartilage which is thought to be important in cartilage structure. In specific embodiments, polypeptides of the invention comprise the sequence: RCKKCTEGPI DLVFVIDGSKSLGEEFVEVVKQF (SEQ ID NO:297); VTGIDSLTSPKARVGL LQYSTQVH (SEQ ID NO:290); TEFTLRNFNSAKDMKKAVAHMKVM (SEQ ID NO:291); GKGSMTGLALKHMFERSFTQEGARP (SEQ ID NO:292); STRVP RAAVFTDGRAQDDVSEWASKAKANGITMYAVGVGKAE (SEQ ID NO:293); BELQELASEPTNKHLYFAEDFTMDISEKLKKGICEAL EDS (SEQ ID NO:294); TORLEBMTQRM (SEQ ID NO:295); PQGCEBQLH (SEQ ID NO:296); and/or YMGKGSMTGLALKHMFERSFT (SEQ ID NO:289). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in placenta, infant brain, prostate, fetal lung and to a lesser extent in endometrium and fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormal placenta and pregnancy, disorder and injury in brain, prostate, and vasculature. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproduction, neuronal, and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, brain and other tissue of the nervous system, prostate, lung and endometrium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to cartilage matrix protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis,

treatment, and cure of abnormalities in placenta and pregnancy, disorder and injury in brain, prostate, and vasculature.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 55

The translation product of this gene is the human ortholog of bovine and hamster CII-3, a succinate-ubiquinone oxidoreductase complex II membrane-intrinsic subunit, which is thought to be important in mitochondrial electron transport chain during metabolism. In specific embodiments, the polypeptides of the invention comprise MAALLLRHVGRHCLRAHSPQLCIRNAVPLGTTAKEBERFVNKNIG SNRPLSPHITYS (SEQ ID NO:298); VFPLMYHTWNGIRHLMWDLGKGLKPQL YQSG (SEQ ID NO:299); MAALLLRHVGRHCLRAH (SEQ ID NO:300); VKSLCL GPALHTAKFAL (SEQ ID NO:301); VFPLMYHTWNGIRHLMWDLGKGL (SEQ ID NO:302).

This gene is expressed in 8-week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolism disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely], e.g., immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, treatment, and cure of metabolism disorders.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 56

This gene is expressed primarily in umbilical vein endothelial cells, human ovarian tumor cells, human meningioma cells, and human Jurkat membrane bound polysomes. In specific embodiments, polypeptides of the invention comprise the amino acid sequence: RVWDVRFAPAKERCYKIQGNV (SEQ ID NO:303); HNFENKLL

RCSWSPDGSKIAAGSADRFVYV (SEQ ID NO:304); and/or WDTTSRRILYKLPG HAGSINEVAFHPDEPI (SEQ ID NO:305). Polynucleotides encoding these polypeptides are also encompassed by the invention.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, immune and cardiovascular disorders and urogenital neoplasias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, neurological, urogenital, reproductive system and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, cells, endothelial cells, ovary and other reproductive tissue, meningima, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:143 as residues: Phe-71 to Arg-76, Pro-82 to His-87, Glu-103 to Ala-111.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. In addition, expression in ovarian tumor cells suggests that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of ovarian tumors, and other tumors and neoplasias. Further, endothelial cell expression suggests a role in cardiovascular or respiratory/pulmonary disorders or infections (athisma, pulmonary edema, pneumonia).

FEATURES OF PROTEIN ENCODED BY GENE NO: 57

The translation product of this gene shares sequence homology with type I collagen. In specific embodiments, the polypeptides of the invention comprise the sequence: GRIPAPSPVPAGPDSR (SEQ ID NO:309); VRGRTVLRPGLDAEPE LSPE (SEQ ID NO:306); EQRVLERKLKKERKKEERQ (SEQ ID NO:307); ARRS

AELAWDYLCRWAQKHKNWRFQKTRQTWLLHMYDSKVPDEHFSTLLAYLE GLQGR (SEQ ID NO:255); and/or RLREAGLV AQHPP (SEQ ID NO:308).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in epididymus, prostate cell line (LNCAP),

5 and pituitary gland; and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormalities of the epididymus, prostate (especially prostate cancer),

10 and pituitary gland. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system and neuroendocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., epididymus and other reproductive tissue, prostate, and pituitary gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to type I collagen, indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of the epididymus, prostate (especially prostate cancer), and pituitary gland.

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

This gene is expressed primarily in the frontal cortex of the brain from a schizophrenic individual.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of

51

the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of schizophrenia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

10 The polypeptide encoded by Gene 59 is homologous to human surface 4 integral membrane protein. In specific embodiments, the polypeptides of the invention comprise the sequence: TGCVLVLSRNFQYACFGLFGIALQTIAYSYLWDLKF LMRN (SEQ ID NO:310); SRSEKSMFAGVPTMRESSPKQYMQLGGRVLLV LMEFTLLHFDASFFSIVQNIYG (SEQ ID NO:311); GTAEFDADQFLRYTKQYLP HVARCLISTFLEDGIRMFQWSEQRDYIDTTWNCGYLLAS (SEQ ID NO:312); LMRNERS (SEQ ID NO:314); ASFLSRTSWGTA (SEQ ID NO:315); and/or ASFLSRTSWGTA (SEQ ID NO:313). Polynucleotides encoding these polypeptides are also encompassed by the invention.

20 This gene is expressed primarily in Hodgkin's lymphoma and lung, and to a lesser extent in many other human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, Hodgkin's lymphoma, tumors or other abnormalities of the lung.

25 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., lymphoid tissue, and pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:183 as residues: Met-20 to Tyr-27.

35 The tissue distribution indicates that polynucleotides and polypeptides

52

corresponding to this gene are useful for diagnosis and treatment of Hodgkin's lymphoma, tumors or other abnormalities of the lung.

FEATURES OF PROTEIN ENCODED BY GENE NO: 60

5 This gene is expressed primarily in bone cancer and stomach cancer, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone cancer and stomach cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, and the stomach, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., bone, and stomach, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of bone cancer and stomach cancer and possibly other cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 61

25 This gene is expressed primarily in epididymus, and lymph node of breast cancer, and to a lesser extent in many other tissues.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormalities of the epididymus, and breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the epididymus and breast, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., epididymus and other reproductive tissue, lymphoid tissue, and mammary tissue, and cancerous and wounded tissues) or bodily fluids

(e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:185 as residues: Arg-57 to Ser-65.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of the epididymus, and breast cancer.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 62

The translation product of this gene appears to be the human homolog of bovine NADH dehydrogenase which is thought to be important in cellular metabolism. In specific embodiments, the polypeptides of the invention comprise the amino acid sequence: SMSALTRLASFARYGGRLFRSGCARTAGDGGVRRHAGGGVHIEPRY
15 RQFFQLTRSQVFQSEFFSGLMWFILWRFWHDSEVLGHFFPYDPFSQWTDDEL
GIPDDED (SEQ ID NO:323), or fragments thereof. Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed in larynx tumor, lymph node, brain amygdala, human cardiomyopathy, and retina.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases affecting cellular metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
25 of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., larynx, lymphoid tissue, brain and other tissue of the nervous system, heart and cardiovascular tissue, and retina, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:208 as residues: Pro-27 to Gln-32, Arg-
35 42 to Glu-51.

The tissue distribution and homology to NADH dehydrogenase indicates that

polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases involving cellular metabolism.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 63

This gene is expressed primarily in amygdala, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormalities of the amygdala. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the amygdala, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., amygdala, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids
15 (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:187 as residues: Gln-17 to Glu-29, Pro-41 to Phe-46,
20 Ser-59 to Ile-70, Thr-97 to Leu-105.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of amygdala.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 64

This gene is expressed primarily in female bladder, and to a lesser extent in chronic synovitis and hemangiopericytoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bladder cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
30 particularly of the urinary tract, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., bladder, synovial tissue, and

55

vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:188 as residues: Pro-2 to Gln-7, Pro-27 to Phe-34.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatments of defects of the urinary tract, especially bladder cancer.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 65

This gene is expressed primarily in fetal spleen, and to a lesser extent in hemangiopericytoma, thymus, and synovial sarcoma.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, defects of immune or hematopoietic systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., spleen, vascular tissue, thymus, blood cells, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The protein product of this gene is useful for treatment of defects of the immune or hematopoietic systems, because of the gene's expression in thymus and spleen.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 66

This gene is expressed primarily in human pituitary and to a lesser extent in placenta and fetal lung.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

56

not limited to, endocrine growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., pituitary and other endocrine tissue, placenta, and pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:190 as residues: Val-38 to Asn-44, Gly-53 to Ser-65.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of growth disorders related to pituitary dysfunction.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 67

The translation product of this gene shares sequence homology with a *Ceenorhabditis elegans* gene of unknown function. In specific embodiments, the polypeptides of the invention comprise the sequence: DPRPNKVLRYKPPSE CNPALLDDPTP (SEQ ID NO:317); DYNMLLGMTFSMCOQLMLKLKWCWVA VYCS (SEQ ID NO:318); FISFANSRSSEDTKQMMSSF (SEQ ID NO:316); and/or MLSSAVVVMSTYLQNPQPMITPPW (SEQ ID NO:319). Polynucleotides encoding these polypeptides are also encompassed by the invention.

25 This gene is expressed primarily in primary breast cancer and lymph node breast cancer and to a lesser extent in adult brain, lung cancer, colon cancer, epithelioid sarcoma, and Caco-2 cell line.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer and tumor tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., mammary tissue, lymphoid tissue, brain and other tissue of the nervous system, lung, colon, and

epithelium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:191 as residues: Asn-34 to Lys-42.

The tissue distribution in a variety of cancer tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of a variety of cancer and tumor types.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 68

The translation product of this gene shares sequence homology with steroid membrane binding protein. The translation product of this gene has recently been published as progesterone binding protein. See Genbank AJ002030. Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

15 AAGDGDVKGTLGSGSESSNDGSESPGDAGAAAXGGGWAAAAALLTG
GGE (SEQ ID NO:320).

This gene is expressed primarily in breast, and to a lesser extent in placenta and fetal tissue.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer or developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of breast or fetal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., mammary tissue, placenta, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:192 as residues: Pro-43 to Asp-49, Gln-54 to Pro-64, Asp-110 to Asp-118, Lys-138 to Tyr-143, Pro-150 to Asp-170.

35 The tissue distribution and homology to steroid membrane binding protein and to progesterone binding protein indicates that the protein products of this gene are

useful for treatment of breast cancers, especially those caused by estrogen and progesterone binding.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 69

Preferred polypeptides encoded by this gene comprise the following amino acid sequence: AADNYGIPRACRNSARSYGAAWLLXPAGSSRVEPTQDISDQLGG QDVPFRNLSLLVVGVGAVFSLFHGTRRRRRPHAXEPGEHTPLAPATAQPL LLWKHWLREXAFYQVGLYMTTRLJVNLSQTYMAMYLTYSLHLPKKFIATPLV MYLSGFLSSFLMKPINKCIGRN (SEQ ID NO:321).

10 This gene is expressed primarily in macrophage (GM-CSF treated), and to a lesser extent in monocytes and dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., macrophages and other blood cells, and dendritic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution indicates that the protein products of this gene are useful for treatment of infection or inflammation or other events or defects involving the immune system.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 70

This gene is expressed primarily in adult brain and to a lesser extent in thyroid, 12 week old early stage human, and stromal cell TF274.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological or neuro-endocrine diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous or endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, thyroid, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 194 as residues: Pro-65 to Cys-71.

The tissue distribution indicates that the protein products of this gene are useful for treatment and diagnosis of neurological diseases or metabolic conditions involving the neuro-endocrine system.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene is expressed in T-cell helper and to a lesser extent in adult brain and adult testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders, meningitis or reproductive problems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, neural and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, brain and other tissue of the nervous system, testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 195 as residues: Val-18 to Tyr-24, Ala-89 to Asp-99, Asp-104 to Ala-117, Leu-121 to Pro-136.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis immune and

reproductive disorders.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 72

The translated polypeptide of this contig has a high degree of identity with the Ob Receptor-Associated Protein deposited as GenBank Accession No. 2266638. No function has been determined for the Ob Receptor-Associated Protein, however it is expressed upon stimulation of the Ob Receptor by Leptin.

This gene is expressed in T-cells and to a lesser extent in endothelial and bone marrow cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, acute lymphoblastic leukemia, hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, endothelial cells, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 196 as residues: Ser-61 to Trp-70.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of leukemia and other disorders of the primary immune system. In addition, since this gene appears to be related to the Ob Receptor-Related Protein, it is likely that this polypeptide is also involved in the Ob/Leptin signal transduction cascade. As a result, this protein may be of use in the molecular diagnosis and therapeutic intervention of obesity and related disorders.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 73

The translation product of this contig has homology with furin, a protein thought to be a key endopeptidase in the constitutive secretory pathway. The identification and initial characterization of Furin was reported by Takahashi and

colleagues (Biochem Biophys Res Commun 1993 Sep 15;195(2):1019-1026).

This gene is expressed in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system such as allergies, wound healing and antigen recognition. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neutrophils and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of allergies or other immune disorders since neutrophils are an important part of an allergic response. Further, since this protein appears to be related to Furin, it can be used diagnostically and therapeutically to treat secretory protein processing disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 74

This gene is expressed in the frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, of the motor activity and sensory functions that involve the central nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of neural disorders that affect cognitive functions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 75

The translation product of this gene shares sequence homology with inorganic pyrophosphatase which is thought to be important in the catalysis the hydrolysis of diphosphate bonds, chiefly in nucleoside di- and triphosphates and essential enzymes that are important for controlling the cellular levels of inorganic pyrophosphate (PPi). The bovine homolog of this gene has been identified by Yang and Wensel (*J. Biol. Chem.* 267:24641-24647 (1992)).

This gene is expressed in osteoclastoma cells and to a lesser extent in epithelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoporosis and other bone weakening diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, and epithelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:199 as residues: Lys-22 to Tyr-28, Asp-64 to Lys-77, Pro-86 to Ile-91, Gln-99 to Pro-119, Tyr-169 to Asp-174, Lys-176 to Gly-181, Trp-189 to Asn-202, Lys-233 to Gly-239, Ser-250 to Asp-257.

The tissue distribution and homology to inorganic pyrophosphatase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of osteoporosis through the removal of bone by demineralization.

FEATURES OF PROTEIN ENCODED BY GENE NO: 76

The translation product of this gene shares exact sequence homology with ATP sulfurylase/APS kinase (GenBank Accession No. 2673862) which is thought to be important in biosynthesis of the activated sulfate donor, adenosine 3'-phosphate 5'-phosphosulfate, involves the sequential action of two enzyme activities: ATP sulfurylase, which catalyzes the formation of adenosine 5'-phosphosulfate (APS) from ATP and free sulfate, and APS kinase, which subsequently phosphorylates APS to produce adenosine 3'-phosphate 5'-phosphosulfate.

This gene is expressed in osteoclastoma cells and to a lesser extent in developmental tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, antibiotic resistant bacterial infections, osteoarthritis and other autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or skeletal structure expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., bone, and developmental tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:200 as residues: Asn-15 to Trp-20, Ser-36 to Gly-41, Pro-103 to Val-110, Pro-134 to Arg-143, Leu-173 to Arg-178, Ser-190 to Ala-197, His-314 to Arg-319, Arg-354 to Asn-362, Asp-391 to Arg-397, Glu-402 to Asp-409, Asp-434 to Leu-439, Glu-441 to Arg-446, Gly-455 to Asp-462, Pro-528 to His-541, Asn-566 to Arg-571, Tyr-574 to Glu-581, Thr-589 to Glu-603.

The tissue distribution and homology to ATP sulfurylase/APS kinase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment or detection of autoimmune diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 77

This polypeptide is identical to the SLP-76-associated protein reported by Musci and colleagues (J. Biol. Chem. 272 (18), 11674-11677 (1997)) and to the Fyb protein

reported by da Silva and coworkers (Proc. Natl. Acad. Sci. U.S.A. (1997). In press). These proteins have been reported to be novel T-cell proteins which bind FYN and SLP-76 and regulate IL-2 production. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: RITDNEGKWLGRARGSYGYIK TTAVEIXYDSLKLKDSLGA PSRIEDDQEVYDDVAEQDDISSHSQSGSGGIFPP PPDDIYDGIIEEDADGFPAPPKQLDMGDEVYDDVTSDFRVSSAEMSQGTNV GKAKTEBKDLKKLKKQXKEXKDPFRKKPKYDGEIRVLYSTKYVTTSTSSKKWGT RDLQYKPGSELVVIQTDDTKVLCRNEEGKYGVLRSYLADNDGEIYDIADGC IYDND (SEQ ID NO:322).

This gene is expressed in CD34 positive cells (hematopoietic progenitor cells) and to a lesser extent in adult spleen derived from a chronic lymphocytic leukemia patient.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, chronic lymphocytic leukemia; hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., T-cells and other blood cells, bone marrow, hematopoietic cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Further, nucleic acids and polypeptides of the present invention are useful both diagnostically and therapeutically in the intervention of immune and other disorders in which the ability to alter IL-2 expression is desired. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:201 as residues: Ala-17 to Lys-37, Val-39 to Ser-45, Lys-59 to His-70, Arg-90 to Leu-95, Lys-97 to Lys-107, Ser-117 to Leu-124, Phe-133 to Ser-138, Trp-146 to Leu-167, Pro-175 to Asn-185, Lys-190 to Ser-211, Pro-213 to Ser-222, His-230 to Pro-235, Pro-240 to Pro-246, Pro-253 to Gly-261, Leu-271 to Leu-303, Leu-305 to Leu-326, Lys-343 to Leu-349, Thr-363 to Leu-371, Arg-373 to Tyr-381, Tyr-391 to Leu-401, Pro-404 to Val-414, Ser-426 to Ser-432, Ile-448 to Ser-457, Gln-462 to Trp-468, Lys-477 to Ser-501, Asp-518 to Ser-523, Ala-541 to Gln-554.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for the treatment of a variety of hematopoietic disorders. The noted expression of this gene in the hematopoietic progenitor cell

compartment - as determined by its expression on CD34 positive hematopoietic stem and progenitor cells - indicates that it plays a critical role in the expansion or proliferation of hematopoietic stem/progenitor cells, as well as in the differentiation of the various blood cell lineages. Thus it could be useful in the reconstitution of the hematopoietic system of patients with leukemias and other hematopoietic diseases.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is homologous to heparin cofactor II (HClII) which is a 66-kDa plasma glycoprotein that inhibits thrombin rapidly in the presence of dermatan sulfate or heparin.

This gene is expressed in apoptotic and anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, thrombopenia T-cell lymphomas; Hodgkin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system - most notably the T-cell compartment, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The homology to heparin cofactor II (HClII) and the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic disorders particularly in thrombopoiesis, most notably of the T-cell compartment. This could include immune modulation, inflammation, immune surveillance, graft rejection, and autoimmunity.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 79

The translation product of this gene shares sequence homology with a mouse

protein believed to represent an integral membrane protein.

This gene is expressed in fetal cochlea and epididymus and to a lesser extent in adult spleen and osteoclastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoclastoma; disorders of the inner ear; male fertility disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inner ear; male reproductive tract; bone; and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cochlea, epididymus and other reproductive tissue, spleen, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:203 as residues: Lys-13 to Gly-23, Cys-38 to Asp-43, Gly-48 to Trp-53, Cys-223 to Ile-237, Ile-240 to Ser-246.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of hearing and fertility disorders. Likewise, it may have a role in the modulation of immune function and in the treatment of osteoporosis.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 80

The translation product of this gene shares sequence homology with reticulocalbin which is thought to be important in the binding of calcium, particularly within the endoplasmic reticulum.

This gene is expressed in endothelial cells and stromal cells and to a lesser extent in osteoblasts, osteoclasts, and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoporosis; osteoclastomas; T-cell lymphomas; Hodgkin's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculature, bone, and immune systems - particularly the T-cell compartments, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, stromal cells, bone, T-cells and other blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:204 as residues: Lys-20 to Arg-27, Pro-32 to Asp-48, Leu-64 to Arg-72, Asp-108 to Lys-114, Gln-128 to Thr-133, Asp-139 to Phe-147, Thr-196 to Ala-204, Tyr-218 to Gln-228, Val-230 to Gln-236, Arg-241 to Lys-255, Gln-276 to Lys-287.

The tissue distribution and homology to reticulocalbin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of bone disorders such as osteoporosis; the diagnosis and treatment of T-cell lymphomas and Hodgkin's lymphoma; and the treatment of diseases and defects of the vasculature, such as vascular leak syndrome and aberrant angiogenesis that accompanies tumor growth.

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

The translation product of this gene shares sequence homology with a family of peptide transport genes - particularly the APTTR2-B gene from *Arabidopsis* - which are thought to be important in the uptake of small peptides.

This gene is expressed in a number of fetal tissues, most notably lung, brain, cochlea, and liver/spleen, and to a lesser extent in osteoclastoma and endometrial tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoclastoma, endometrial tumors, cancer, leukemias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone and endometrium, expression of this gene at significantly higher or lower levels may be

routinely detected in certain tissues (e.g., fetal tissue, pulmonary tissue, bone, brain and other tissue of the nervous system, cochlea, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:205 as residues: Lys-186 to Asn-199, Pro-202 to Ala-207.

The tissue distribution and homology to peptide transport genes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the control of cell proliferation, owing to its strong expression in fetal tissues undergoing active cell division, as well as its expression in a variety of tumors or cancers of adult tissues. Potentially, it may regulate the uptake of peptides that stimulate cell proliferation. This gene product may also be useful in stimulating the uptake of a variety of peptide-based drug compounds.

FEATURES OF PROTEIN ENCODED BY GENE NO: 82

This gene is expressed in fetal liver and spleen and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and tumors of a hematopoietic and/or endothelial cell origin; leukemias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and/or vasculature, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, endothelial cells, vascular tissue, and tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:206 as residues: Met-1 to Asp-9, Arg-66 to Gly-76, Asp-164 to Arg-171.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for the treatment of disorders of the immune system. Expression of this gene product in both fetal liver/spleen and endothelial cells indicates that it may be expressed in the hemangioblast, the progenitor cell for both the immune system and the vasculature. Thus, it is most likely expressed in hematopoietic stem cells, and may be useful for the expansion of hematopoietic stem and progenitor cells in conjunction with cancer treatment for a variety of leukemias.

FEATURES OF PROTEIN ENCODED BY GENE NO: 84

The translation product of this gene shares sequence homology with NADH dehydrogenase which is thought to be important in cellular metabolism.

This gene is expressed in fetal dura mater and to a lesser extent in T-cells and hypothalamus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases affecting cellular metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, T-cells and other blood cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:208 as residues: Pro-27 to Gln-32, Arg-42 to Glu-51.

The tissue distribution and homology to NADH dehydrogenase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases involving cellular metabolism.

FEATURES OF PROTEIN ENCODED BY GENE NO: 85

The translation product of this gene shares sequence homology with I-TRAF, a novel TNF receptor associated factor (TRAF)-interacting protein that regulates TNF receptor-mediated signal transduction. This protein is thought to be important in regulating the cellular response to tumor necrosis factor (TNF), which is an important mediator of inflammation.

This gene is expressed in endothelial cells and to a lesser extent in glioblastoma and osteoblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation; glioblastoma and osteoblastoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, bone, and glial cells and tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:209 as residues: Glu-15 to Thr-22, Glu-46 to Leu-62, Arg-103 to Glu-119, Gln-127 to Glu-132, Asn-152 to Trp-158, Gln-191 to Gln-210, Glu-264 to Thr-271, Tyr-282 to Leu-288, Trp-319 to Thr-331, Glu-335 to Ser-348, Ser-353 to Ser-358, Asp-382 to Asn-392.

The tissue distribution and homology to I-TRAF indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory diseases, including rheumatoid arthritis, sepsis, inflammatory bowel disease, and psoriasis, particularly where tumor necrosis factor is known to be involved.

FEATURES OF PROTEIN ENCODED BY GENE NO: 86

This gene has homology with a candidate gene involved in X-linked Retinopathy reported by Wong and colleagues (Genomics 15:467-471 (1993)).

This gene is expressed in a T-cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and autoimmune diseases; T-cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory disorders such as sepsis, inflammatory bowel disease, psoriasis, and rheumatoid arthritis as well as autoimmune disease such as lupus. It could also be useful in immune modulation and in the process of immune surveillance. The present invention can be used diagnostically and therapeutically to treat X-linked Retinopathy.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 87

This gene is expressed in human brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain disorders; neurodegenerative disorders; tumors of a brain origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:211 as residues: Cys-32 to Tyr-38.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of CNS disorders such as epilepsy, paranoia, depression, Alzheimer's disease, and schizophrenia. It could be useful in the survival and/or proliferation of neurons and could effect neuronal regeneration.

73

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total Seq. NT	5' NT 3' NT Clone Seq.	5' NT Start Codon	5' NT First AA of SEQ NO: Y	First AA of Last AA	First AA of Secreted of First AA	First AA of Last AA	First AA of Last AA
1	HAGEW82	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	11	1679	247	1607	353	125	1		30
2	HAGFY16	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	12	1830	87	1786	128	126	1	26	44
2	HBMCF37	xxxxx 03/19/98	pbLscript	98	1487	79	1487	170	212	1	44	69
2	HFLQB16	209641 02/25/98	Uni-ZAP XR	99	1653	394	1637	413	213	1	25	81
3	HALA60	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	13	1212	1	1212	99	127	1	24	38
4	HAPBL78	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	14	2061	882	2061	900	128	1	22	22
5	HASAV70	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	15	1412	10	733	103	129	1	20	109

74

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total Seq. NT	5' NT 3' NT Clone Seq.	5' NT Start Codon	5' NT First AA of SEQ NO: Y	First AA of Last AA	First AA of Secreted of First AA	First AA of Last AA	First AA of Last AA
6	HBNAF22	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	16	1052	276	880	538	130	1	17	62
7	HBNBL77	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	17	683	1	683	181	131	1		29
8	HCDDR90	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	18	1054	86	1007	86	132	1	23	52
9	HCEEF50	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	19	1393	132	1393	192	133	1	17	56
10	HCEMU42	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	20	1215	277	1070	401	134	1	18	215
11	HCENE16	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	21	2042	614	2011	793	135	1	26	48

75

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
12	HMSJJ74	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	22	1872	21	1872	69	69	136	1	23	24	67
13	HCUBF15	97923 03/07/97 209071 05/22/97	ZAP Express	23	289	1	289	89	89	137	1	29	30	51
14	HE2DE47	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	24	3533	2821	3532	808	808	138	1	30	31	539
14	HE2DE47	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	100	1145	435	1115	515	515	214	1	22	23	80
15	HKMLH01	209179 07/24/97	pBluescript	25	1148	171	907	196	196	139	1	26	27	56
15	HE6DG34	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	101	734	25	734	295	295	215	1	36	37	48
16	HE9DG49	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	26	717	1	717	70	70	140	1	27	28	200

76

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
16	HE9DG49	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	102	713	17	713	78	78	216	1	28	29	202
17	HELBA06	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	27	1099	1	1099	38	38	141	1	22	23	215
17	HELBA06	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	103	1080	1	1080	149	149	217	1	25	26	185
18	HSLFM29	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	28	941	171	941	128	128	142	1	42	43	101
19	HELBW38	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	29	756	62	756	294	294	143	1	30	31	111
20	HETHN28	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	30	2100	408	2093	496	496	144	1			19

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	5' NT First AA of Signal Pep	AA SEQ NO: Y	First AA of Sig Pep	First AA of Sec Portion	Last AA of ORF
21	HFCDK17	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	31	1448	475	1392	567	567	145	1		29
22	HFEAF41	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	32	456	1	409	21	21	146	1	28	98
23	HFKFL13	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	33	1326	1	1322	210	210	147	1		7
24	HFSBG13	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	34	710	1	710	242	242	148	1	16	38
25	HFTBE43	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	35	1188	110	1161	178	178	149	1	26	130
26	HFTDJ36	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	36	956	1	938	144	144	150	1	21	31
27	HKTACT7	97924 03/07/97	Uni-ZAP XR	37	1603	974	1581	1104	1104	151	1		13

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	5' NT First AA of Signal Pep	AA SEQ NO: Y	First AA of Sig Pep	First AA of Sec Portion	Last AA of ORF
28	HLHSH36	97924 03/07/97	pBluescript	38	1089	55	1067		209	152	1		7
29	HLHSV96	97924 03/07/97	pBluescript	39	629	1	629	119	119	153	1	32	67
30	HLQBQ86	97924 03/07/97	Lambda ZAP II	40	1964	408	1793	581	581	154	1		25
31	HLTBX31	97924 03/07/97	Uni-ZAP XR	41	1522	13	1123	126	126	155	1	32	194
32	HLTCJ63	97924 03/07/97	Uni-ZAP XR	42	875	1	875	43	43	156	1	18	90
33	HMKAH44	97924 03/07/97	pSport1	43	843	1	843	171	171	157	1	30	30
34	HMQAJ64	97924 03/07/97	Uni-ZAP XR	44	489	3	489	55	55	158	1	19	89
34	HMQAJ64	97924 03/07/97	Uni-ZAP XR	104	489	6	489	58	58	218	1	22	89
35	HOABG65	97924 03/07/97	Uni-ZAP XR	45	534	1	534	17	17	159	1	18	88
36	HODCL36	97924 03/07/97	Uni-ZAP XR	46	1374	1	1374	15	15	160	1	20	173
36	HODCL36	97924 03/07/97	Uni-ZAP XR	105	640	58	640	72	72	219	1	20	137
36	HODCL36	97924 03/07/97	Uni-ZAP XR	106	1529	40	1399	54	54	220	1	27	47
37	HODCL50	97924 03/07/97	Uni-ZAP XR	47	596	1	596	269	269	161	1	27	44

79

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
38	HODCV74	97924 03/07/97	Uni-ZAP XR	48	851	99	822	170	170	162	1			22
39	HODCZ16	97924 03/07/97	Uni-ZAP XR	49	2020	569	2020	638	638	163	1	17	18	69
40	HTOEU03	97924 03/07/97	Uni-ZAP XR	50	2432	848	2432	99	99	164	1	19	20	322
40	HTOEU03	97924 03/07/97	Uni-ZAP XR	107	2435	849	2435	928	928	221	1	31	32	69
41	HPBCJ74	97924 03/07/97	pBluescript SK-	51	2340	1627	2340	150	150	165	1	60	61	319
41	HPBCJ74	97924 03/07/97	pBluescript SK-	108	805	92	791	239	239	222	1	21	22	82
42	HPMBU33	97924 03/07/97	Uni-ZAP XR	52	601	188	601	432	432	166	1			30
43	HSAUL66	97924 03/07/97	Uni-ZAP XR	53	359	1	337	142	142	167	1	18	19	71
44	HSIDQ18	97924 03/07/97	Uni-ZAP XR	54	1141	1	1141	25	25	168	1	30	31	280
44	HSIDQ18	97924 03/07/97	Uni-ZAP XR	109	1166	21	1166	433	433	223	1	30	31	42
45	HSJBB37	97924 03/07/97	Uni-ZAP XR	55	1560	63	1148	217	217	169	1			22
46	HSJBQ79	97924 03/07/97	Uni-ZAP XR	56	1507	164	608	57	57	170	1	19	20	326
46	HSJBQ79	97924 03/07/97	Uni-ZAP XR	110	586	4	586	35	35	224	1	23	24	183

80

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
47	HTEGA76	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	57	450	1	450	83	83	171	1	35	36	68
48	HTEJN13	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	58	1147	1	1147	163	163	172	1	15	16	158
48	HTEJN13	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	111	1134	1	1134	155	155	225	1	19	20	70
49	HTHBL86	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	59	777	1	777	115	115	173	1	18	19	122
50	HTSFO71	97958 03/13/97 209072 05/22/97	pBluescript	60	1191	48	598	52	52	174	1	30	31	128
50	HTSFO71	97958 03/13/97 209072 05/22/97	pBluescript	112	1333	594	1333	829	829	226	1			9
51	HAPNO80	209235 09/04/97	Uni-ZAP XR	61	1580	443	1554	114	114	175	1	1	2	371

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
63	HHGDB72	97958 03/13/97 209072 05/22/97	Lambda ZAP II	73	620	1	620	96	96	187	1	18	19	131
64	HHGDI71	97958 03/13/97 209072 05/22/97	Lambda ZAP II	74	581	156	581	248	248	188	1	32	33	68
65	HHSDI45	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	75	1843	537	1786	630	630	189	1	27	28	44
66	HHSEB66	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	76	1441	116	800	167	167	190	1	36	37	64
67	HJPAZ83	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	114	1076	398	1076		575	228	1	11	12	22
68	HLDBO49	97958 03/13/97 209072 05/22/97	pCMVSPORT 3.0	78	2776	18	1888	187	187	192	1	14	15	169

83

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
69	HLDBQ19	97958 03/13/97 209072 05/22/97	pCMVSPORT 3.0	79	1525	401	1480	534	534	193	1	22	23	65
69	HLDBQ19	209226 08/28/97	pCMVSPORT 3.0	115	1487	401	1487	534	534	229	1	22	23	131
70	HMSGT42	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	80	1563	33	1077	40	40	194	1	32	33	91
71	HMWIC78	97957 03/13/97 209073 05/22/97	Uni-Zap XR	81	1020	18	780	238	238	195	1	23	24	175
72	HTTCT79	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	82	770	101	770	286	286	196	1	26	27	69
73	HNGJU84	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	83	481	1	481	58	58	197	1	20	21	24
74	HNTAC73	97957 03/13/97 209073 05/22/97	pCMVSPORT 3.0	84	644	1	623	14	14	198	1	25	26	72

84

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT 3' NT Clone of Seq.	5' NT Start Codon	5' NT First AA of Signal Pep	AA SEQ NO: Y	First AA of AA Sig Pep	Last AA of AA Sig Pep	First AA of AA Secreted of Portion	Last AA of AA Secreted of ORF
75	HOSEI45	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	85	1351	435	1284	98	98	199	1	12	288
75	HOSEI45	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	116	1350	428	1283	545	230	1			27
76	HOSFD58	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	86	2527	290	1747	56	200	1	30	31	623
76	HOSFD58	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	117	2527	288	1747	477	231	1	32	33	60
77	HSAUM95	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	87	2566	1843	2566	251	201	1	30	31	648
77	HSAUM95	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	118	1098	375	1098	677	232	1	21	22	28

85

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT 3' NT Clone of Seq.	5' NT Start Codon	5' NT First AA of Signal Pep	AA SEQ NO: Y	First AA of AA Sig Pep	Last AA of AA Sig Pep	First AA of AA Secreted of Portion	Last AA of AA Secreted of ORF
78	HSAUR67	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	88	540	1	540	83	202	1	32	33	54
79	HSKD181	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	89	1863	152	1165	188	203	1	11	12	265
79	HSKD181	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	119	1679	152	1166	315	233	1			17
80	HSKDW91	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	90	2478	1149	2449	92	204	1	19	20	314
81	HTLEX50	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	91	2058	476	2058	414	205	1	20	21	206
82	HSKHL65	97957 03/13/97 209073 05/22/97	pBluescript	92	1411	345	1411	157	206	1	69	70	194

98

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
82	HSKHL65	97957 03/13/97 209073 05/22/97	pBluescript	121	1411	345	1411	526	526	235	1	37	38	71
83	HHFGA11	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	93	2187	147	2184	397	397	207	1	30	31	329
83	HHFGA11	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	122	2256	138	2063	228	228	236	1	19	20	95
84	HWTBL40	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	94	757	524	608	445	445	208	1	20	21	57
85	HBXFG80	97957 03/13/97 209073 05/22/97	ZAP Express	95	2394	481	2394	523	523	209	1	1	2	391
86	HCACY32	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	96	672	1	672	117	117	210	1	21	22	25

87

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
87	HCEDO21	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	97	1419	1	1419	207	207	211	1	20	21	37

88

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988).

Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeech, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeech and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brulag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity.

For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are

considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

95

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C-termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

96

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobei et al., *J. Biotechnology* 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (*J. Biol. Chem.* 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., *Science* 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln; replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, or 701 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any

combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Polypeptide fragments of SEQ ID NO: Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Segments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et

al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting).

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab)₂ fragments) which are capable of specifically binding to protein. Fab and F(ab)₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAb or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Elton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

35

Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells. The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and lac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium

phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

30 Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat

polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO. X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO. X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and prescreening by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library).) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined.

First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying

personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

5 A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

15 In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

20 A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 mCi of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

5 Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

15 Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

20 At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

25 The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

35 A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the

proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect

interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

- For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

- 10 Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thyroid, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

- 15 Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstrom's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

- 25 A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

- 30 Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes

- Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, broncholitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

- 15 Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., *Corynebacterium*, *Mycobacterium*, *Nocardia*), *Aspergillus*, *Bacillaceae* (e.g., *Anthrax*, *Clostridium*), *Bacteroidaceae*, *Blasomycosis*, *Bordetella*, *Borrelia*, *Brucellosis*, *Candidiasis*, *Campylobacter*, *Coccidioidomycosis*, *Cryptococcosis*, *Dermatocycoses*, *Enterobacteriaceae* (*Klebsiella*, *Salmonella*, *Serratia*, *Yersinia*), *Erysipelothrix*, *Helicobacter*, *Legionellosis*, *Leptospirosis*, *Listeria*, *Mycoplasmatales*, *Neisseriaceae* (e.g., *Acinetobacter*, *Gonorrhea*, *Menigococcal*), *Pasteurellaceae* Infections (e.g., *Actinobacillus*, *Haemophilus*, *Pasteurella*), *Pseudomonas*, *Rickettsiaceae*, *Chlamydiaceae*, *Syphilis*, and *Staphylococcal*. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections.
- 35 A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997)). The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue

regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit

(antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of

positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type

Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone. A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method

comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO: Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO: Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO: Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO: Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lafrind BA	plafind BA
pSport1	pSport1
pcMVSPort 2.0	pcMVSPort 2.0
pcMVSPort 3.0	pcMVSPort 3.0
PCR [®] 2.1	PCR [®] 2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128,256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988), Ailing-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Ailing-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the fl ori of replication ("ori"), such that in one orientation, single stranded rescue initiated from the fl ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pcMVSPort 2.0 and pcMVSPort 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain

5 DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., *Focus* 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR^{2.1}, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., *Nuc. Acids Res.* 16:9677-9686 (1988) and Mead, D. et al., *Bio/Technology* 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

10 The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

15 Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO.X.

20 Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported.

25 The oligonucleotide is labeled, for instance, with ³²P- γ -ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above.

30 The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 μ l of reaction mixture with 0.5 μ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM MgCl₂, 0.01% (w/w) gelatin, 20 μ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

5 Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., *Nucleic Acids Res.* 21(7):1683-1684 (1993).)

10 Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

15 This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

20 This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is

used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

5

Example 2: Isolation of Genomic Clones Corresponding to a

Polynucleotide

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X, according to the method described in Example 1. (See also, Sambrook.)

10

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P³² using the rediprimer™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

20

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

25

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions: 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on

35

either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

15

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis. Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.₆₀₀) of between 0.4 and 0.6. IPTG (isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

20

25

30

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high

35

affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl, pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl, pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After

renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate, pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHEAa. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an *E. coli* origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgamo sequence, and 6) the lactose operon repressor gene (*lacIq*). The origin of replication (*oriC*) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEA by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filttron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commae blue stained 16% SDS-PAGE gel when 5 μ g of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polydrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcMD1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-35 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five μ g of a plasmid containing the polynucleotide is co-transfected with 1.0 μ g of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One μ g of BaculoGold™ virus DNA and 5 μ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50 μ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μ l Lipofectin plus 90 μ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life

Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.)

5 After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 μ l of Grace's medium and the suspension containing the recombinant baculovirus is used to infect SF9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, SF9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 μ Ci of ³⁵S-methionine and 5 μ Ci ³⁵S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell.

25 A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIV and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used

include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., BioTechnology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphatases by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five µg of the expression plasmid pC6 is cotransfected with 0.5 µg of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 µM, 2 µM, 5 µM, 10 nM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion

proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fe portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fe portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fe portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

20 Human IgG Fe region:
GGGATCCGGAGGCCAAATCTTCTGCAAAACTCACACATGCCCCACCGTGCC
CAGCACCTGAATTTCGAGGGTGACCCGTCAGTCTTCTTCCCCCAAAACC
CAAGGACACCCCTCATGATCTCCCGGACTCTCTGAGGTACATGCGTGGTGT
GGACGTAAGCCACGGAAGACCCCTGAGGTCAAGTTCAACTGTAAGTGGACG
CCGTGAGGTGCATATGCGCAAGACAAAGCCCGGGAGGAGGACGTAACAAC
AGCAGTACCGTGTGTGTCAGCGTCTCAACCGTCTGCAACGAGCACTGGGCTG
AATGGCAAGGAGTACAAGTGCAGGTCTCCAAACAAGCCCTCCCAACCCCC
ATCGAGAAAACCATCTCCAAAGCCAAAGGGGAGCCCGGAGAACACAGGT
GTACACCCCTGCCCCCATCTCCGGGATGAGCTGACCAAGAAACCAAGTCAAGCT
GACCTGCCTGTGTCAAAAGGCTTCTATCAAGCGACATCGCCGTGGAGTGGGA
GAGCAATGGGAGCCCGAGAACACTACAAGACACACGCTCCCGTGTCTGG
ACTCCGACGGCTCTTCTTCTCTACAGCAAGCTCAACCGTGGACAGAAGACA
GGTGGCAGCAGGGGAACGCTTCTTCATAGTCTCCGTGATGATGAGGCTCTGC
ACAACCACTACACGCGAAGAGGCTCTCCCTGTCTCCGGGTAAATGAGTGC
35 GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide.

Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')₂ and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x 10⁵ cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 μ l Lipofectamine (18324-012 Gibco/BRL) and 5ml OptiMem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2 μ g of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50 μ l of the Lipofectamine/OptiMem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150 μ l OptiMem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200 μ l of DNALipofectamine/OptiMem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl₂ (anhyd); 0.00130 mg/L CuSO₄·5H₂O; 0.050 mg/L of Fe(NO₃)₃·9H₂O; 0.417 mg/L of FeSO₄·7H₂O; 311.80 mg/L of KCl; 28.64 mg/L of MgCl₂; 48.84 mg/L of MgSO₄; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO₃; 62.50 mg/L of NaH₂PO₄·H₂O; 71.02 mg/L of Na₂HPO₄; 4320 mg/L of ZnSO₄·7H₂O; .002 mg/L of Arachidonic Acid; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitic Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCl; 7.50 mg/ml of L-Asparagine-H₂O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCl-H₂O; 31.29 mg/ml of L-Cystine-2HCl; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCl-H₂O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCl; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na·2H₂O; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of

Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of L-Inositol; 3.02 mg/L of Nicotinamide; 3.00 mg/L of Pyridoxal HCl; 0.031 mg/L of Pyridoxine HCl; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCl; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B₁₂; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipic Acid; 0.081 mg/L of Sodium Putrescine-2HCl; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20mM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 μ l for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300 μ l multichannel pipetter, aliquot 600 μ l in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in

many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN- α , IFN- γ , and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

Ligand	TyK2	JAKs Jak1	Jak2	Jak3	STATs	GAS(elements) or ISRE
IFN family						
IFN- α /B	+	+	-	-	1,2,3	ISRE
IFN- γ	+	+	+	-	1	GAS (IRF1>Ly6>IFP)
IL-10	+	?	?	-	1,3	
gp130 family						
IL-6 (Pleiotrohic)	+	+	+	?	1,3	GAS (IRF1>Ly6>IFP)
IL-11 (Pleiotrohic)	?	+	?	?	1,3	
OnM (Pleiotrohic)	?	+	+	?	1,3	
LIF (Pleiotrohic)	?	+	+	?	1,3	
CNTF (Pleiotrohic)	-/+	+	+	?	1,3	
G-CSF (Pleiotrohic)	?	+	+	?	1,3	
IL-12 (Pleiotrohic)	+	-	+	+	1,3	
gC family						
IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >> Ly6) (IgH)
IL-7 (lymphocytes)	-	+	-	+	5	GAS
IL-9 (lymphocytes)	-	+	-	+	5	GAS
IL-13 (lymphocyte)	-	+	?	?	6	GAS
IL-15	?	+	?	+	5	GAS
gp140 family						
IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
IL-5 (myeloid)	-	-	+	-	5	GAS
GM-CSF (myeloid)	-	-	+	-	5	GAS
Growth hormone family						
GH	?	-	+	-	5	
PRL	?	+/-	+	-	1,3,5	
EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
Receptor Tyrosine Kinases						
EGF	?	+	+	-	1,3	GAS (IRF1)
PDGF	?	+	+	-	1,3	
CSF-1	?	+	+	-	1,3	GAS (not IRF1)

143

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994)), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GGGCGCTCGAGATTTCGCCGAATCTAGATTTCGCCGAATGATTTCCCGG
10 AAATGATTTCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTGGCAAGCCTAGGC:3' (SEQ ID NO:4)

15 PCR amplification is performed using the SV40 promoter template present in the B-gal promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTTCGCCGAATCTAGATTTCGCCGAATGATTTCCCGAAATG
20 ATTTCGCCGAATATCTGCCATCTCAATTAGTCAGCAACATAGTCCGCC
CTAATCTCGGCCCATCTCCGCCCTAACTCCGCCAGTCCGCCATTTCCGC
CCCATGCGTGAATAATTTTATTATTATGACAGAGGCCGAGCCGCTCGGC
CTCTGAGCTATTCAGAACTAGTGAAGGCGCTTTTGGAGGCGCTTAGGCTTT
25 TGCAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

144

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomycyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

25 Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/Neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final concentration of 10^7 cells/ml. Then add 1ml of 1×10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Genticin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100, 000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophane covers) and stored at 20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2×10^7 U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM MgCl_2 , and 675 uM CaCl_2 . Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting 1×10^6 cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of 5×10^5 cells/ml. Plate 200 ul cells per well in the 96-well plate (or 1×10^5 cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and BGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:
 5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)
 5' GCGAAGCTTCGCGACTCCCGGATCCGCTC-3' (SEQ ID NO:7)
 Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as 5×10^5 cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to 1×10^5 cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ml of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF- κ B (Nuclear Factor κ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- κ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- κ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κ B is retained in the cytoplasm with I- κ B (inhibitor κ B). However, upon stimulation, I- κ B is phosphorylated and degraded, causing NF- κ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- κ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- κ B would be useful in treating

diseases. For example, inhibitors of NF- κ B could be used to treat those diseases related to the acute or chronic activation of NF- κ B, such as rheumatoid arthritis.

To construct a vector containing the NF- κ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- κ B binding site (GGGGACTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:

5'-CGGGCTCGAGGGGACTTCCGGGGACTTCCGGGGACTTCCGGGGACTTCCGGGGAC
TTTCCATCTGCCATCTCAATTAG-3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5'-CGGGCAAGCTTTTGGCAAAGCTAGGC-3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)

Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5'-CTCGAGGGGACTTCCGGGGGACTTCCGGGGGACTTCCGGGGACTTCC
ATCTGCCATCTCAATTAGTACGAACCATAGTCCGCCCTTAATCCGCCCA
TCCCGCCCTAATCCGCCAGTTCGCCCATTTCTCCGCCCATGGCTGACT
AATTTTATTATGACAGGCCGAGGCCCTCGGCTCTGAGCTATTC
CAGAAAGTAGTGGAGGCTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:
3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2- promoter plasmid (Clontech) with this NF- κ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- κ B/SV40/SEAP

cassette is removed from the above NF- κ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF- κ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- κ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μ l of 2.5x dilution buffer into Optiplates containing 35 μ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μ l Assay Buffer and incubate at room

temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

151

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in

Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO₂ incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

152

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO₂ incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10⁶ cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10⁶ cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca⁺⁺ concentration.

Example 19: High-Throughput Screening Assay Identifying Tyrosine

Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na₃VO₄, 2 mM Na₄P₂O₇ and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 40°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 40°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg₂+ (5mM ATP/50mM MgCl₂), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl₂, 5 mM MnCl₂, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or complement to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

155

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1µg/ml) for 2 hr at room temp. (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprottyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 µl of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1µg/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

25

Example 21: Method of Determining Alterations in a Gene

Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30

seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies).

156

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenin-deoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human col-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

25

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 µg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide.

Next, 50 μ l of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate.

Add 75 μ l of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 μ g/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μ g/kg/hour to about 50 μ g/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile.

Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

161

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

5 pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

15 The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the tier of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

35

162

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

5 The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

163

(1) GENERAL INFORMATION:

- (i) APPLICANT: Human Genome Sciences, Inc. et al.
(ii) TITLE OF INVENTION: 87 Human Secreted Proteins
(iii) NUMBER OF SEQUENCES: 323
(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.
(B) STREET: 9410 Key West Avenue
(C) CITY: Rockville
(D) STATE: Maryland
(E) COUNTRY: USA
(F) ZIP: 20850

15 (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.44b storage
(B) COMPUTER: HP Vectra 486/33
(C) OPERATING SYSTEM: MSDOS version 6.2
(D) SOFTWARE: ASCII Text

25 (vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
(B) FILING DATE: March 19, 1998
(C) CLASSIFICATION:

30 (vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER:
(B) FILING DATE:

35 (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: A. Anders Brookes
(B) REGISTRATION NUMBER: 36,373
(C) REFERENCE/DOCKET NUMBER: P2004PCT

40 (ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (301) 309-8504
(B) TELEFAX: (301) 309-8439

45 (2) INFORMATION FOR SEQ ID NO: 1:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 733 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

164

GGAGTCCGGA GCGGAAATCT TGTGACAAA CTCACACAG CCCACCTCC CACACACTG 60
AATTCAGAG TGCACCTCA GTCTCTCTT TCCCCCGAA ACCGAGGAC ACCCTCATGA 120
TCTCCCCAG TCTTGAGTC ACATCCCTG TGTGACACT AACCCAGAA GACCTCTAG 180
TCAGTTCAA CTCGTACCTG GACGCTGAG AGGTGCTTAA TCCGAGACA AAGCCCCGAG 240
AGAGCCAGTA CACACACAG TACCTGTGG TACAGCTCT CACCTCTCT CACGAGACT 300
GCGTAAATG CAGAGAGTAC AATGACAGG TCTTCACAA ACCCTTCCA ACCCTCATG 360
AGAAACCAT CTCGAAAGCC AAGGCGCAC CCGAGAAAC ACAGTGTAC ACCCTGCCC 420
CATCCCGGGA TGAAGTGAAC AAGAACAGG TACAGCTGAC CTCCTCTGTC AAGAGCTCT 480
ATCCAGACCA CATTCCCTTG GAGTGGAGAA GCATTGGCCA GCGGAGAAC AACTACAGA 540
CGAGGCTTCC GTAGCTGAG TCCGACGAGT CATTCTCTT CTACAGCAG CTCACCTGG 600
ACAGAGCAG GTGCGACAG GGAAGCTCT TCTCATGCTC GATGATGAT GAGCTTCTC 660
ACAGACCTTA CAGCAGAG ACCCTTCCC TGTCTCCGG TAAATGAGTG CGACGCGCC 720
GACTCTAGG GAT 733

30 (2) INFORMATION FOR SEQ ID NO: 2:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser
1 5

45 (2) INFORMATION FOR SEQ ID NO: 3:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGCGCTGAG AATTCCTCCA AATCATGAT TCCCGAAT GATTTCCTG AATGATTTT 60
CCGAAATAT CTCGATCTC AATTAG 86

165

166

(2) INFORMATION FOR SEQ ID NO: 4:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGCGCAAGCT TTTTCGAAAG CCTAGAC

27

15

(2) INFORMATION FOR SEQ ID NO: 5:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATTT CCGCGAATTC TAGATTTCCT CGAATGATTT TCCCGGAAT GATTTCCCG
AATATCTGC CATCTCAATT AATCAGAAC CATAGTCCG CCGCTAACTC CCGCCATCCC
GCCCTAACT CCGCCCAAGTT CCGCCCTATC TCCCGCCCAT GCGTGACTTA TTTTITTTAT
TTATCGAGAG GCGAGGCCG CTTCCGCTTC TGACTATTC CAGAATGACT GAGGAGCTT
TTTTCGAGGC CTAGCGCTTT GCAAAAGCT T

60

120

180

240

271

(2) INFORMATION FOR SEQ ID NO: 6:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CGCGTCGAGG GATCGACAGG ATAGACCCC GG

32

(2) INFORMATION FOR SEQ ID NO: 7:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

55

60

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCCAGACTTC GCGACTCCCG GGATCCGCTT C

31

(2) INFORMATION FOR SEQ ID NO: 8:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGGAGCTTTC CC

12

(2) INFORMATION FOR SEQ ID NO: 9:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CGCGCCCTCA GGGAGCTTTC CCGGGGACTT TCCGGGACT TTCGGGACT TTCATCTCG
CCATCTCAT TAG

60

73

(2) INFORMATION FOR SEQ ID NO: 10:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

45

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGCGA CTTTCCCGG GACTTTCGG GACTTTCGG GACTTTCGA TCTGCATCT
CAATATGACA GCACCATAG TCCCGCCCTT ACTTCGCC ATCCCGCCC TACTTCGGC
CAGTTCGCC CATCTCCGC CCATGGCTG ACTANTTTT TTTATTTATG CAGAGGCCGA
GGCGCCCTCG GCGTCTGAGC TATTCAGAA GTATGAGGA GCGTTTTTTC GAGGCTTAG

60

120

180

240

167

CTTTCGAAA AACCTT

256

(2) INFORMATION FOR SEQ ID NO: 11:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1679 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

15 GAGAGCCAGC CCGGCGATCG CTTTCAGCAT GCGGACAGCG TACCCATCTT TACCTACGTC
AAGAAAAATG CCGGCTGCGC CAGCTCCGTC ACCGGTACG CTTTCGAAA AGCGATGAG
20 AAGAGCTGCC TACGCGACGA CTCGTGCGAG TCCCTGAAAG ACCGCTACCT CAGACCACTG
CGGGGCGAGG AACATTAATTA CCGCTGCGGG GAGCGCGCGG TGAACCCCTC CTCGCGAG
25 CTCACACGGA ACGCGAGAGA GAGCCCGGAG GCGCGCGATTA GCGGGGAACT ACAGAAATAG
AGAACTCCAG ATTTCCTCTA AGAAGATAT CTGAAAGAG AGATCCAGCA GAATGAAAGA
GCAATCAAAA AGATGCTTGT GGAAGCCACC CCGGAGTTTG AGGAGTTTGT GGTGAGTAGAG
30 AGCCCTCTCG ATTTCGAAT ACATTAATCT ATGTGTGATG ATGATCCACC CACACCTGAG
GAAAGCTCAG AAGACAGCC TGAATGAGAG GAAAGAGAG AGAAGAAAAA AGTTTCTCAA
CGAAGGTGAG GAGTGCAT TAAAGTCAAT CCGCAATTAA TGAAGAAATT TAACTTGGAT
35 CTATCAAGCG TTACAGAGCC CTTTCTAAAA AATTAATGTG AGCTGAGAGC TACTTCCGCC
TTCTTACGCT CTGTGCGAGG AGCTGATGCA TATCCATTTT GGTCCCGACA AGATGACATA
GATTTCCAAA AAGATGATGA GGAATACAGA GAGGCAATGG TCAAAAAATT TGTGTCTCAG
40 AATGTAGCTC GAGAGATTTA ATTTCGAAAG AATTAATTTG CAGATTAATG AGAAAAAGAA
AAGATCAGTG TAGTGTAGCT GATTAAAAAA AATTGTACCC AATGATCTTT AGAGAGTTCT
45 TGCATTGAAA CTGGCACTTA TTTTCTGACC ATGCGTGTG TTGCTGTGTG AGTCTGAGAT
TTTTGTAGCC AAGCAAGATT GTTACAGGGG ATAAAAAGAA AAGAAATTTG ATGTATTTAC
50 AGCTGTCTCT GAAACAGTAT CAAATGTCTT ATGAAAGAAA GATCTAAATC AGACAGAGAT
TGTCTTAAGT AGTATTAATC CATTGTGAGA ATGGAACCTT TGCATATGTA GTGACAAAGT
GAAAGAAAT TTGAGAGAGA TAGGCAATTT CAGGCAAGAT AAGTAATCTC CTCTCTTTG
55 GCAAGAGCTC CTTTGAATTG GGAATAGTTC CAATTAAGGA ATCTAGAAAT AGGAGAGAT
TTAATTATGA GGCCTGAGAC ACGGATTAAT CCGAAGCCCT TGTCAATTCC CCGAGTAGGC
60 TGTGATTTCT AGACTGCTTT GAAATAGCTG TATTCATTTT GCTTACTTAG TATTTGAGTA 1380

168

CCCTGCTCTT TGGCTGTCTT TTTTTCGAG CCGTTTCAG TCAAGTCTCG CCGATGCTTT 1440

TCTTAACTTA CCGCTCAAGT TTCTTAAAA CCGGACAGA AGCTAGAGA GTTAAAGAA 1500

5 TAAATTAAT TGTATTAAT GTTATTAAT GATTAATTT TGTGTAAGC ATGTGTAGG 1560

10 ATTAAAAA TTATGTGANT GATCCAGTT TGTGTGTG TTTTCAATCT TCAAAATAAA 1620

TATTAATTTG TATTCGAAA AAAAAAAA AAAAAATCTG CAGAGCAGCA ACGGAAATTC 1679

(2) INFORMATION FOR SEQ ID NO: 12:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1810 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

15 GCGACCGGCG CTTTCAGCTA GCTGCTGCG TGGCTGCTCT TCCCTCTGCG CCGGTCGCGA
TGGCTTACG CTTGCGGCG CTGCGCGCGG TCGAGCAGCC TCGGAGGCG GTACAGACAG
20 TTGCGAATG AAGAGAGCTC TCGAGAACT GAAAGAGTTC CAGGTGATGC TCCCTCACT
TACAGACGA TTCTGCGAGA GAGCGCAAT TATTTGACT ACAGAGATGA GTCTGCTGTT
25 GCGAAGCCCC CAGCTTAAGA TGTAGCTACA ACATCCCA GTTATGATGA ACGGAGAGAG
CGAAGGCTG AAGCTACAT CCGTTGAT CTGCGAGAG ATGAGAAATT TGTGGTCCG
30 GATGATTTG ATGATCTCTA CCACTGAGG ATAGAGAAAG ATGGAATTTT CATTGTACCT
TTTTCATGCG CATTCTCTTT TAACTGANT GGGTTTTCC TGTCTTTTTC CCGACCACT
35 TGTCTGCGAG GAGGTGATG GCGCAATCA GGAATTTGAT TCTCTTAAT TAAATGAGTC
CTGATTGTCA GATTTTCACA CTAATTTCCCT GGAATTTTTC ATGTGCACTA CTGCGCTGCG
40 TGAAGTTCAG GATTTTCAGG GCGCAATCA GGAATTTGAT TCTCTTAAT TAAATGAGTC
45 TGGGTGTC TTGTTTTCAG CTTTCTGCG TTCTGCGAG GATTTTACA TTATGCAAAA
GTTCGAGAGA TCGCAGAAC TTTCGAAAT CTCCCGAGA CCGAGATCTT CTTTATTTAT
50 TAAAGATTT TTCTGCGAAA GCGCTCTGCG GATTTTACA TTCTCTCTA AGAAGAGAGA
GAAAGCTCG AAGAGATGA TCAAGATGCA GAAAGAGAG GAAATTAATC CTGCTTTTAA
AAGATTAAT ACTGTGAAA AAGTCAATTC TCTCTATTTG TTCTGAGTGC TAAATTTTAA
55 ATATTAATG CAGATTTCTG TAAATCTTA ATCATTAATG GTTAAATTTT GAAAGAGCTC
TTGCAATCA GTCTGTAGAG TATTAATAT CCGTTAATTA TTGTTTGGAG TCAATTTAG
60 TACATGAGC GATGCTGCG TAGTGTGAGG GGGGCAAGCT TGGTTATATC ATCTCCATC 1080

169

5 1140 TCAAAATGAA CTGTGAATTA ATATGTGTAA GATATGTATA ATGTGTGCGA TTTHAAGCG
1200 GTTCTTCTGA AGCTTAAGCT TTCTGTATCA CTGTGTATTT GCACATATATC CATATTTGCT
1260 GTTCAAGTTA ATCTAGAAAT TTATTCATTT CTGTATGAAC ACCTGGAGCC AAAAAATATG
1320 TGCAAAATA CATTAAAGT GTGTGCAAAA ATAAATCTTT ATTTGTATTA TATTAAGCAT
1380 TAATTTTATA TAGCTGTAT TTCAATCTT GCGTACCTT ATTTGACCTA AGGATTTCTA
1440 AAGGTCTGTG CACTGTATTA AACAAGAGC ACTAGGATAC AATGAAAGCT TAATTAATAA
1500 AATGTAAATC TTGACACTCT TTCTATATTT AGCTTCTTC ACCCCCAACC CCACCCCAAC
1560 CCCCCTTAT TTCCCTTTGT CTCTGTGTGA TTAGGCCAAA GTCTGGAGCT AAGGAGAGA
1620 TTAGTACTT AGGAGCAAG AAGAGATAG CTGTGAACCT TTGATATGAT CCTTAACATA
1680 CTGTACTACT TCTTTTATCA ATGTGTATAG AGAAACCACT GGTATATAT GTAGATATGAT
1740 GTGCTTTCTG CCCAAGTGT ATATCATCTT GGTTCCTAT GTTAAAACTG TAATATCAAC
1800 AGACATTTAA TAATATATCT TTGTGTAGCA CTTTTHAAA AAAAAAATA AAAAAAATA
1830 AAAAAAATA AATCCCGCG GCGGCGCCCN

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1212 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 13:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

40 TGTTCAGAT TGTACTTTT GTTACAGCA AAGTTTGTG TAGTGTGAG TAGTACGCTC 60
TAGACTGATC TTTTCTTAAA TCAGAAAGTG ATTAAGTAT GCACAAACAA AGCGAGTTT 120
TTCTTTTTC TTTATTCAG ACTATTTAT TAGATCAA CTCTGTCCA GCGAGTTAC 180
TAGCTCTAC ATACTGTCTG ACATGATAT AGSTTAAT ACTTTTAAA TTATATCAA 240
ATACTCAAT GTAGATATTT CTTAAGTTCA ATAGCATTA ACTAGATAA TCGTTTATG 300
TTATTTTAT TTCTCTGTCA TAGAATTTCA ACTTTGTACC ATCTTAAAC TAGTTTCTTA 360
TAAAAATAG AGGATAGAT CATTAAGTTT TATCCAGTTT TAAAACTGG AAGGAAAGG 420
TAGAGCTCT CCATTTAAA ATAGTTGAT TCGTTTATTT TTACATATTT AGTGCATTC 480
GTATATCAAC TGGCCCTCAA TGAACATTT AAGTCTCTG AATTTTACTA AACTGACTTT 540
TTTCAACTT TGGAGATTT TTGAGGGAG TGTGTAAT TTCCAAACAC TCACCTCTTA 600
CTCAAACTT CAATTAAT ACATTTTC AAGGAGAGC ACCTTTTATA TTGTATAGT 660

170

5 720 TTTCATATA AACCTTATA TACAGTCAAC AAGAGGTTG TCTGTCTATG GTTTAGCAAA
780 CATTGTCTT TCTTTTGA AGTGTATTT CATTGTGCA ACAGAAAGTG AGAAACACT
840 GCGAGGTTG ATCTACTCT GAGGTAGTTT TTACAACTA CCATTTCCC TCCATGAAT
900 TATGTAAAT TTATTTATC TTGTGAAA GTTGAGAGA TAGTAAAGA ATTAGAAT
960 TAAATTAACA GGGAAAAATA TGTAAAGTAA AAGCAATAAA TATTTGTTC ACTTTGCTAT
1020 CAGATGTTT ACTATCAGAT ATTTATATA TGGAGCAT TTATATTTT ATTCATTTCC
1080 CATTATAGA CCGATAAA TATTTTGA TCAGACATTT GCGTTTGT TA TGTGCAATTA
1140 AATGTCTTT TGTACTGTAA GTTACTGTTA ATTTGAATAT TTATTTGAAC TGTCTCCCTG
1200 TCCCTTATA ATATTAAGTT GTTCTAGAA CTTTATATCA TCTTATATA GAATACTTTA
1212 AGAAAAAAA AA

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2061 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 14:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

35 GTTTTCTC CACTTCCGG ACHTCTCCT GCGAGTGGG CAGAGTGGG TCMAAGGCA 60
CCAGTCTCG CTCGGTCTC TGGGATCG GACCGCGCG GCGCGCCCG AGCGGATGT 120
TCCGGGCTT GAGCAGTTG TTGGGCTTC AGCAGCCGTT GCGAGGCGGT GGGCAGCCA 180
ATCGAGTGC TCACCGGAG CAGCGTTCG AGACGTTGC TGAATCTCG GAGGAGGAC 240
TCCAGCAGC GCGAGACAG GAGCTCTCC ACCAGGCCA AGACTTCGC ACTATTTAT 300
TTAATTTTC ATCTCTCC ACNAAAAA TAATCTGATC AGTTGCTGAA ACAGCAGAA 360
CAATAAGAA ATCCGTGAA GAAGAAAAA TAGATGGCAT CATTGACAG ACATTTATG 420
GAGTTTTC GAAGAGAG AAAAAATTG TTGAGAGCA ACATACAGAG AAGTCAGAG 480
CAGCTGTCC CCATGGGTT GACATAAG ATGAAGAAC AATTCACAA CAATTTTGG 540
CCTTATAGC TGAAGAGG AATTTCTTC GTGACCTCC GCTTGGGTT CAATTTAAT 600
TGAAGTTGA TGAATGATC CCGTGGGCC TGTCTATCT CAGAGAGAT GAGCTGCTAR 660
CAGATGAGA TTTCCTCG TTCTTAAT TGTGAGGA GAGTGTCTT GAGGAGACTA 720
CTTTTACGC GTCTCTG TTAAGCAGT AGCGGCGTC AGCGGCGTC CTGCCAACA 780

171

840 GAGAGCCGCG GAGAGGAG GAGAGAGCA ATGCGAGCA GGAAGATTG CCGTCGAGA
900 GCGAGTACAG CCGAAACCG CACCCCTTCT AATCGAATCT GAGCTTAAA CTCAGAGGGA
960 TCGAGAGGAA ATTCTTACTA GCGCAGGTCT TTCTGAGTTT CTCAGTAGTG CCGTCAGTCC
1020 CTGTACCTTA AATCGAGAG ATTCTAGGAA AGAATGAGG GAGCTAGTGC TTGACAAAAA
1080 GCGAGAGGAG ACAGCCGTCAG TCGAGAGAGA TTCTCGAGAT TCGGAAAAAG AACTCGAGCA
1140 GGAATCTTAA GAAATGAGAG TCGTCGACAG ATCTGAAAA GAGATGAAA ACTGGAGTAA
1200 GGAATGAGAG AAAATGCTTC AAGAGAAAA TTAGCTGTTC CTGAAATGAG AGAATATTC
1260 TTACAGCTCT GCAATCTAGC ATTAAATCT AGATGTGAC AATTACTGAA TCGAGAGGCA
1320 TGAAGAGTAA TAAATTTTAG AAATTCAAAA TTATCTTTTT TTCAAGTTGA AACCTGCTTC
1380 TTCTACTTTA AAAAAGTATA TGAACAGCTT ACTTCTAATA ATCGAAAAAG GATGTTTTAT
1440 AGAACATTC TTTAATATTA AGTTAGAGAT GTCTGATAG GAGATATGCG TATCTTTGCG
1500 AGGAAAGAT AAGTAAATTT TTAGAGTCTT GTTTCGATAG AGTTCAAAA TAAATTTTTAT
1560 TCCGAGTACA TGGTTTTCTA AATATCTGTA CTCGACATTC GATTTTAAAT GATATGAGGG
1620 TGTAAAGTAA CTAATCTAAT GGGTTGATTA CTATCAAAAT GAGCAATTTA TACGAAAGAA
1680 CTTAGAGAGA AGACTTTCA GAACTATTCA GTTCCAGGCT AATTCTTAAA ATTCGACTTG
1740 AAAGCGAAAA GATTAATATAC ATTAGTTGGA TTTTATATAT ATTAAGCTCA GACATTTTAA
1800 CATTAAGAAA TACTGTGAG CCGATGCTGG GTGCTGACAG CCGTATATCC CAGCATTTTG
1860 GAGAGCCGAG GTGGGAGAT CAGCGAGGCT CAGGAGTTGG AGACCAAGCT TCCGACATTA
1920 GTGAAGAGCT GTCTTTACTA AAAATACAAA AATTAGCCGG GAGTGTGCG AGGCAAGCTT
1980 AATCCAGACT ACTAGAGAGG CTTTGAACG CAGAGAGGAG AGGTTGAGAG GAGCTGAGAT
2040 CCGGCGACTG CAGTCGAGCG TGGGTGATAG AGTGAATTC AGTCTCAAAA AAAAAAAAAA
2061 AAAAAAAAAA AATGACTTGG A

45 (2) INFORMATION FOR SEQ ID NO: 15:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1412 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

60 CCGTTCATCT GCGTTGCGAG GAACCTGTCT AGCGAAGACT TCTGAGGCCC CAGCTTGCCC
120 AGAAGAGCTCT GTAGAGGTGC TCGTAGTAGC CCAAGATCTT CCAATGTGCT CCGTGTGTCT

172

180 CTGTGTGTCG CCGTCTGTCT CAGTCTCTTT GTACTGTGCG TATTTCTTGG GTTCTGAGAG
240 AGAGAGAGAC AAGAGAGATA CATTAAGAG AAGAGAGAG TCGAGCTTGG TCGGAAATCT
300 CTAAGAGTAT GCGCCCAATC TCGAGAGAGC ACAGATAGAG ACAAGATCCC TCAAGCTAAT
360 AGAAGATCCC TAAAGAGAGA TCGACAGAAAT AGGTTTACT CCACTGTGGA AATAGCGAAA
420 AAGATGAAA ATCCCACTAC ACTGTCTAGC ATGCCAGCA CACCAAGCTT AATTGCGTAT
480 GAAATGTTAA TCGAGAGAGC AGTGCAGTCC CTAATGTCT TCGTCAAAAA AAAAGCAATT
540 CTGCGGCCAA AGAAGAGAT CAGAGAAAT CAGTGAATTT ACTGATAGCA TCAAGAGAGA
600 ATGAGAGAG TTGACTTTTT TCGAGAGTAA ATTAATCTGAG ATGCTCTTT AGATTTAGAA
660 GTTCATTAAT CAGTCACTAG CTGAGAAATC TCGTCAGCC CAGAGGTTT AATGATCTCA
720 TCCCAAAAT GGAATGTGTA ATGTCAAGAA ACCATTAATA AGGTGCTTAG AAGTATCTCT
780 AATAAATCT AAATGCAAG TCAAGAGAT TAAATGAGC CTGTTATAT AATGATGCT
840 CAGGTGAT GTTGTAGATT TCAATTGATC CAGAGGCTTG GATGTGAGAG TTATACAGAG
900 AGTGTGCTTA CAGAGAGGCG AAGAGAGCA AAACAGAGAG ACAATGTCCAG CAGAGAGAGA
960 TCGACGTAGC AAAATATGAT GTATTAATTT GCTCTAATA CTATGTGCC ACAGTTATGC
1020 TGAAGCTTCA CTAAATGTGC AGACATGCTG TGTGCCCTCA TGAATGTGCG TCCAAATAGAA
1080 TGAAGTACTT TCAATGAGAG TTGTAGCAGG CTGAGACAGA GATTTCCAGA GGGCAGGTG
1140 TGAATCCAGA GAACTTGAAG GTCAAGATTC ACAAGATGTA AGAATGAGGG TTAGCTGACAA
1200 TGTTTGAGAG ATATATATAT GAGAGAGAG AAGTGTGAT GCGCCAGAGA CAGAGAGCTC
1260 CAGCGAGCT TCAATTTAGC AGTGTGTGCG AAAAGAAAA TGTAGGTTTT AAGCTGTGCG
1320 CAGAACCAT CCGAATTAAG AGACGAGATC TGAAGTACA TTGTAATATCT AGGTAGAGAG
1380 ACTTGAGATC AGCGAGTAGAG ACTGTGAGGG CAGCGGGGCG AATGGATAT GTAAAGCTTT
1412 TAAAGATGAT TAAATCTTCA TTAGTGTTTT TT

45 (2) INFORMATION FOR SEQ ID NO: 16:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1052 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

60 TTCTGTCTCT CTCTCAATCC CTCTGTGTCT TCGTGTGCTCT GTCTGTGTCT CTCTGTGTCT

173

120	TCCTCTCCGC TCCTCTCTCT TCCTCTCTGC TCCTCTCTCT TCCTCTCTCT TCCTCTCTCT
180	CCCTCTATCTC TTCCCTCTCT CTATCTCTCTC CTCTCTCTCTC TCCTCTCTCTC TCCTCTCTCTC
240	CTCTCTCTCTT CTCTCTCTCTC TCCTCTCTCTG CCTCTCTCTG CTTCTCAGCTT CCGTCTCTCT
300	GTCTGTGCTCC TTCCGAGAA CTCGACGATG AGGGCAGCCT GGGCACTCTA ATGCTCTACTC
360	TGCTGTGTAC ACAAGAAAA TATTGGGGTC ACTGGCGAGC CCACCCACAC TCACGACGAAT
420	CTCCACTCTTA GTCCCTCTAA CAAACAGGCC TTCACTCTCT CTCCACTCTC AGCAATTTCT
480	ATTCTGATGC CATTGGCTTC AGATCAGAGT GTTTTAAATC ATCAGCGCCT GCGTTATCTCC
540	TGCTCTGAGCC AGGCACAGGG GTCTCTCAGT GGTCTCTGTA CCTCTCTCTC TTGAGCGATG
600	TTGCTCTTTTAT TTATTCTACTT TTACTCTCAC CCTCTCTCTG TACCAGGAGG GGCACACTTCA
660	AAGCCAAGST ACAGGGTGAT ACTCTGTGCT CGAGCATATG TTTTCTCTCAC TTCCTTCTCTC
720	CACCTACCCC CAGCAAGGTG CCTGGGAGA CTTCAGCAGA TGTTTCATTT TGGCCTCTGCC
780	AGTGGCTGAA AGCAGGCCCTC CAATTCGACTG TGAACCTCTG CTTTCCCTCAGC AGCTTTTCCCA
840	GAGAGGGAGA GGGGCTCTTC ACAGCCGGG TTCTCTCTCTCT GCGTCTCTCTCC TCGTCTCAGCT
900	GCAGGCATTC TGAGGGGCAA CGTGAGGAA GGGCCAGGGA TGCATGGGAT TTTAATTTGTT
960	TCATCTACGC TTCCCTCTCTG CAAAGAAACA GTCAAGTCTCT CTCAGTGTCT TTCTGCAATTT
1020	CTGTGTGATGG ACGAGGAAT CTTTTCTACG TTTCTAANTA TGTTCACCAA ATAAANAATTC
1052	CAATTTTATAT TTGGGAAAA AAAAAAAAAA AA

(2) INFORMATION FOR SEO ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 683 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

60	AAATTGGCGAG AGGCACTTAT CATGTACATA TAGCCTGTTT TTTHAGCATTT TTAGACAAG
120	TAGGCATNTT CTTTTCGTC CAGAACTCA TACCTTAGTA ATTGTAGTGG GCTGATAGCT
180	CATTGCCAT ACACAAGAT CTACACAAC CTCCTGAATA AACATCCCC TTATTCAGAA
240	ATGCCCTTTC CTATTTCAT ATTGCAGCT TGCTTACAA TTTCGAATCT GTCTTCTGT
300	TTACAGAGA TATACAAAT TCCCTTTGTA TGAATCTCTT ATATCTCTGG ATTTCCTTTT
360	GTGTTTGTCTA CCNAGGCGC TGCACATAGT GAGAAGATTT TGCATGACT GTGAGCTCTA
420	CCACACCTGG AATTAGGAT CACCATATG AGAAAAAAA TTGAGGATAC AATATAGATT

174

480
ATCTATGTGTT ATTGGCGATAT AANTATCAGA TGTWTCNTG ACTTAAAGACC CTGTGCTGATAT
540
WAACCAATG CAGATAAATW TAATAAAATW TWATAAAATW TWATCAAAATA ATGATAGTGC
600
TAATCAAAATA CTTCAAATTT GCACAGGTGAT TTATTTCTTA AATATGTGTA ACACATGTGA
660
GCCATATAC TGAGTGCTACT GGTAAATATA ACAGATTCCTT GCAGAAAAAA AAAAAAATAA
683
ACTCGAGGGG GGCCCGTACC CTT

15 (2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1054 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

25	AAACTCATTT AGGTGTACACT ATAGAGAGGTA CGCCTGCAGGG TACCGGTCCGG GAATTCCTGGG	60
	GTCCGACCCAC GWNCCGGCG ACAAGATGCG AGCAGCTGTGT CGGAGCGTGA AGGGCCTGGT	120
	GGCGGTAAATA ACCGAGAGGAG CCTCTGGGCGCT GGGGCTTGCC ACGGCGGAGCG ACTTGTGTGGG	180
30	CAGGGAGCCT CTGCTGTGTGTT TCTGTGACTTG CCGTACTTCGG GTGGGGAGCG CCAGGCCGAG	240
	AAGTTAGGAA ACAACTGCGT TTTCGGCCCA GCGGAGCTGTA CCTCTGAGAA GGAATGTGCA	300
35	ACAGCTCTTGG CTCTACGAAA AGGAAGTTT GGCCTGTGTGG ATGTAGCTGT CAACTGTGCA	360
	GGCATATCGG TGCTGTAGCA GACTGTAGAC TTAAAGAGGG GCGAGAGCCA TACCTTGGA	420
40	GACTTCCAGC GAGTCTCTGA TGTGAACTCT ATGGGCACT TCAATGTGAT CGGCTGTGTG	480
	GCTGTGTAGA TGGGCCAGAA TGAACACGAC CAGGGAGGCC AAGCTTGGGT CATCATCAAC	540
	ACTGCCATGG TGCTGCTCTT CGAGGGTCAG GTTGAGCAG CTGCATACTC TGCTTCCGAG	600
45	GGGGGAATAG TGGGCATGAC ACTGCCCATTT GCTTGGGATC TGCGTCCCAT AGGTATCCGG	660
	GTGATGACCA TTGCCCCAGG TCTGTATTGG ACCCCACTGC TGACCAGCCT CCCAGAGAAA	720
50	GTGTGTCACT TGTGTGCCAG CCAGATGGCC TTCCCTTAGC GACTCGGTGA CCTTGGTCAG	780
	TATGCTCACC TGTTACAGGC CATCATCGAG AACCCATTCC TCAATGGAGA GGTCAATCCGG	840
	CTGGATGGGG CGATTCTGAT GCAGCCTTGA AGGGAGAGGG CAGAGAAAC ACAGCTCCT	900
55	CTGGCCCTTC TTTCCTCTGG GTACTACTCT CCAGCTTGG AGGAGAGCCA GTAGCCATTT	960
	TGTAACTGCC TACCAATGCG CCTCTGTGCG TAAATAAATC TCTTTTCTC ACANAAAAA	1020
	AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA	1054
60		

175

(2) INFORMATION FOR SEQ ID NO: 19:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1393 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

5 GGAACAACT GGGATATGTC AGCGTTAGC TACTCAATC CTTCAGAAAG GTAAACATC 60
TTACACGGGA CTGAGAAAC AACACATGCT CTTCAGATGTA TTCAATGATC CTTCAGTTTA 120
ATGAGAGACA CCGGAAGGTC AGAGAGACA CCCCCCTCC ACTGTTCCC AACAGAAC 180
TCCCCAGCA GATGCTCTG GTCTATGATC TCTACTGTT TCTTAACTG TGGGCTCTG 240
CCACCCCCA GAGAAATGG AAGCGTTCCA GABAAAGTCA TGGAAACCT GCTCAACTT 300
TTTGGAACTT TTGAGATCAT CTCAATCAATG CGAATCTCA AACCTGGAG AAGATGCCC 360
CTGAAATCC GAGAGTTCA GCAACCGCTA CAGCTCTCT GACCCCGAG GCAACCCAC 420
ATCCCTTAG GCGGCGCAC GGCACAGGAC CACCAAGAG CTGAGCCCT CTGGCCACA 480
GAATCTTCT CTGATCTCCA ATGCGTCCC GTACAAAT CTTCAGACA GCGCGTTGC 540
CAAGCCAAA GCGCTTCCA GAAATCCCC ACTGGCGAG GAAATGAGC TGAATCCAG 600
CACCACTCT GAGATCTTC GCAATGATG GATTTATTC TCTGAGACA GGTCTATCC 660
35 CTCTGCAAC CCGTGGATCC GAGAGGCTG CCAAGCGAG ATGGGAGCC AGGAGAAAG 720
CGCGGTAGG AATGCCCTGC TCTCCCGAA GATGCACT GCAATGGAG TACCGTATG 780
40 TTGCTTGGG TTGCCAGGG GTCTGACA CACAGAGGA TTTCATGGCC ATGAGAGAG 840
CAGGCGCTGT GTATTAATAC CTTCATTTT TAAATCAAG TCCACTGAAA ACCACTTGG 900
TTTTCAGAGT TCTGAGAAC ACCTGGGATC ACAGATGGA ATTGTTCCC CTTCAGAGA 960
45 TTTTATATC ATGTAAACT CTAAATTTAT CTATCTGTA TATACATAAA TCGGTACCC 1020
ATGCTTGA GACCACTTC TACCTTACA CTCTATCTC TCCAGAGAT GGCATATTT 1080
50 TGAAGATGC TAAATGTGT GATGTGATG GCGTGAAG GCTGTAGAC GAGAGTTCCC 1140
TGGGAGAGT CTGTCTTTG GTATGAAAT TTCTCTCTT CTTCGTATG GAAATTTCC 1200
CTTCAGTAC TGAAGTCTC TCGATAGCC ATCGAAGGC TTCCGTAGAG TTCAAGAAAG 1260
TTCTCTTGG CAGAGAGAG TAGCTAAC TATAGATAG TGTCTTATG GATCAATAG 1320
ATGTACCTG TCAATTAAT AATATATAA AACCCACT GGAATGCTG AAAAAAAAA 1380
60 AAAAAAAAA TCG 1393

176

(2) INFORMATION FOR SEQ ID NO: 20:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1215 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5 ACGAAAGTT TTCCAAATG GAAAGCGGC AGTGAAGCA AGCAATTA TGTGATGAG 60
ATCAATCAT AAGACCCCA GCGTTTAC TTTATGCTC CGATGCTAT GTTGTGTCA 120
ATTGTAGCG GATTAAGAT TCAACAGGA AACACTATG ACCATGTA CGCCAGCTT 180
20 TATTAAGCT CACTATAGG AAGCTGTGA CGCTGCAAG TACCGTTCC GAATTCCTG 240
GTCAACCAAC GGTTCGCCC ACGGTCCTT GAAATTTCA ACTGCCCGG AAGATGAGG 300
25 TGAAGTCCC CCGCCCTAGA GGTGCCCTG CGAAGAGAG AGCTAGAG GAAATTTGG 360
AGCGTTTCC TGGCAAAAG ATTTCATCA ACCTCAAGC ATGCGCTTT CTGCCCTCT 420
30 GCGTTGGA TCCAAAGTCA CTCTGCCCC CCATTACCC TATGAGTGA GCGCCCGAG 480
CTGTGTGCA GACAGAGGA AAGACCCCC ATGATGAG CGGCCCCAG TGGTTGTGA 540
ACCAATCTT GATTAAGCT GATATGTT CTGTGGGAC ACGTGTAGA TCCTAGAAG 600
35 CAAAGATCC GGAAGCAG GCAATGAT TCAATTTAT CGCCAGGAA ACTGGGTGT 660
CGTGGAGG CTGACACAC ATTACCGTA CATTGCAAG ACCATGATT ACCGGGAAAC 720
CATGATCCC AATGAAGCC CTTGCTTCA CGCCAGGTC AACTTTGAG ATCTATGGA 780
40 CAGAAATCC ACTGATGAG AATGAGAT TACTGAGCA GAGAGCGAG TACAGCTTC 840
CACAGATCA GGAAGATTA TCCCTAAC CAAATTTCC AAGCTGATG GATCTCTCC 900
45 TGAAGTGG ATTTATGCC CCAAGACAC ATCAATGAA GATCTTAG AAGAACTTA 960
TGTGCCCTG CTAAAGAAC TCCAGAGGA GTATGAGAG GCGATGGGA TCAAGAGAC 1020
50 CCGAAATAC AAGAGCTCT ATTGTATG AGCTGGGCG AACACACTC CTCCCACT 1080
TCTGTCCAG CCTTGAAGC TGAAGCACTT CTTTTCAAG TCCATATAA GAGACCTTA 1140
TGAATCTCC AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 1200
55 AAAAAAAAA GCGCC 1215

(2) INFORMATION FOR SEQ ID NO: 21:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2042 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

10 CTGCATCCAG GCGCAGATTA ACTCGGTAT CTGTGGTCT GAAGAGAGA AATCGAAT 60
 GCAAGGCTT ACTAGAGTC ATCAGAGCA CTATATATC AGTATATGAA AGAGTTGG 120
 AGTCTCTCT TTGATCTTAC TGAGGCTTTT CTTCGAGA AGAGAACTT ACTGACAG 180
 AGAGTCAAA AAGATTGGA AAGTTTATA CTCATTAAT ATATTACTTA GTCAGTCT 240
 ACCAGCTCT GGAATGTTT GAGAGGCTG CTCATATTT CCATAGTACA CTAAAGCC 300
 AGCTTGACA CATGCTTAC CATCTATAG AGTGGCTAT CAATGCTCT ACTTGTAC 360
 AGTTTACAT CATTAAGCTA TCTTTATGG AGGCTAGGA CTGTTATACA GCTCTAAT 420
 TCATTTTGG TCAATCTGA AAGATCTAG CCACAGAGA CACTCTCTGA GCTGAGAG 480
 AAGTGCAGA GCTTTATCT CAAGAAAGG GGAATATAG AAGTGTCTG ATCAATACT 540
 GTTGTACTCT CATGCAAT GCCCACTCT CCATGCAAGA CAACATAGA GAGCTTAT 600
 TTGATTAACA GTCTGACTT AGAGCTTAA GGAATAAGA ACTAGATAG GAGGAAGA 660
 TTGCGAAAA AGCTGTCCAG TTTCGACCG GTGACTGTG TGATGCCAT TCTGACTAG 720
 AAGAGAAAT GAGTACTAG AGACTTTAG ATTTTGAGA ACCAGAGAA CTTTTCTAT 780
 TGGGTGAGA CTATGCTTTT GAGGAAAG AGTTCTTTCA GATTGATGT TATGTACTG 840
 ACCATATGA AGTTGTCCA GACCACAGT CTCGTGTTA GGTCTTGA TTCTTTGAA 900
 CTGACATGA GAGAGCGTGC AAGATCATTA AACGCGAAT AGCCATCTTA GAGCCCTTA 960
 CTGTAGACT GAATGACAG TATTATCTGT TGATCAACAG AGAGTCCAG TTGAAATG 1020
 CAGATGCTTA CTATGATAG ATGATTTGA AGTTCCCTT TGTGACAGG CTAAAGGATC 1080
 CTGATTCACA CATTTGAAA AAAATTAATA ATCTTAATA GTGAGCACT AGTACTACC 1140
 AGCTCTTTT AGACTCTCT AGAGACCCA ATAAAGTAT CCTAGACTT ATAGGGGAG 1200
 ATGTTCTTGG CCTTGGCAT TTAGTATAG TTGAGTTGG CGTCTCTAT GCGAATACA 1260
 TTACTGAGA TCCGAGAA GAGTGGAAA ATTTGGCAC ATCATTTGGA ACATTACAA 1320
 TTATTTTGG ATTACTGTA AAGCATCTT GAGCGCGCC AGGAATAGA AGTTGACTA 1380
 GAACTTAGTA AAGAGTGT TATGTTCTC CGAACAAAA TGGAGAGAT CAGAACAG 1440
 ATGGCCCTGA CTTAATCTT GTTTTAAAG AAGGAATG TCCATATG AAGTATCTT 1500
 TTTCCTTAGT CAGAGAGGCC CAATTCAT GTGATGTTA CTTTATAG CAGTGAATG 1560

5 CAGTTTGAAC TTGATACA GTCACTGAG TGTTTGCTAG GATCTTAAG AACATAAGT 1620
 TAATTAAMA CTACACCTA ATTATGAAA TTGCTTTGT AAAGACATGT GATTGTATTT 1680
 TTAGATGCTT GTTCTCTAT AAATAAGA CATTTTACC CTCAGTTTCT AAATGTAGAC 1740
 TATTGTTGG CTAGTACTTG ATGATTTCT TGTAGAAA AATGCTGGST AATGTACTG 1800
 GTACAGGCC TGTATATA TTAGATTGA AAAGTAAT TCTATAGTTA CTCCTTCTAA 1860
 AATATTGAC TTCTAATTT CCCCCACC AAATCTTTC CTTTGTGAA ATACTAANA 1920
 CTAGTTATG TTATATAAA GTGTAAATG GTTTCTCTTA ATTATAGCAG AAAAGGCTT 1980
 TGTTAGAAT AAATAAAT GACTTATTC ACTAATGAA AAAAAAAAA AAAAAAAAA 2040
 TT 2042

(2) INFORMATION FOR SEQ ID NO: 22:

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1872 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

60 GGTGAGCCC AGCGTCCGA TTGCGCTAGA GCTCTCTGTA CCGAGAGCC CAGCGAACC 60
 120 TGCGGATGAT GTCGGCGAG TTTATCTTT GTTGGCTTT GGTAACTAG TGGTCCCTC 120
 180 AAGCATCTC AGTCTCTCT GTGTTTATG AATCTAGAC AAGGAGTCC TTATAGAGCC 180
 240 AAGCGACAG GAGCGGAAG GACAGTCCC AAGGATGGG GCTGTCTTTA CTGTGCGAA 240
 300 CCGAATATT GCTCTCTCA GCGAACGAG GTTGAACCA CACCACTTT CCGAGACAG 300
 360 TCACTAGCC CTGCGGAGC RGAACCA CA AGGCGAGGA GGTGAGGCC CAGGCAAGTG 360
 420 AAGAGGAT GTCTTGGGT TTTTAAAGTA GTTGAAGCTG ACCTCTCTGA CTGCTTCTC 420
 480 CCGCGGGGG ACTGCAACC GCTCAGGCTT GCGCGAGAG CATGCACTTC CGGTCCCTGC 480
 540 AAGCGGTGAC CTAAAGCTGG TGCACCTTC AGTACCGAG GAGGACTGAC TTGACAGAG 540
 600 AAGGCAAGC CCGATGACA CAGGCTGGA AGATCTGAG CCGCACTCT GTCCCTGCA 600
 660 ACCAAGAT GCATGTGAG TGTGCGAGT CCCACAGCT CACATGCGC TCCCGCGCA 660
 720 ACGCGAGCA CAGGATCTT CAGGACTTC TGACCTCAC AAGTCAAGTG GACCACTTC 720
 780 CACTCACA GATGTGAA CGTCTCTTTA AATGCGATT TTAGAGCTC GCGATGCT 780
 840 GTGCTGCA TCTTTCTAT TATGCTCAG GATAGTTCA TTTCTTGCA CATAGTGAA 840

179

AAAGATTAG CTGAGATAT TTGCTCTTG AATGACGTC ACCGCGATA TAGGATATGC 900
TTGTATCCC CGGTACTCC TCCGCTGTT TTTTAACTT TTCCAGACC TCGCTCATA 960
5 AAGATATTA TAGGATGTC TCTTAAATG TGAACCTGG TGTGCTGCC GGGCGATCT 1020
GGGTGGCTCC ATGAAAGCT GACTCTGCC CCGACCGGGC TTCTTGAGG AGCTGATTC 1080
TCTTATATAT CATCTTATC TCTGGAAAT CTACTGTAA ATCATCTCG TATTTTCTG 1140
10 GAGCTTCA CATGAGTAG AATGTGAT TTCCGCTGTA GGTCTCTTC GTCCCGCTA 1200
TCTGCGGCT GTACCTGCC CACCTGTCA GATTACTGT GTGTGTAG AGGATATAT 1260
15 TTTAAAGAA CTGCGCTGAA AAGTCTTAG AAAGCGATG AAAGCGAGA ACTGTCTT 1320
TAGCGATTT TTCTTTGTA GATGCGAAA GTTTAAAGG GAGAGAGAG TTGTCAATGG 1380
CTGTCTCTTG GGGTTTTAA TCTCTCTAC CTGTAGATA ACCCTCGGC TTGTCTTACC 1440
20 AGCGAGCGA AAAGCTCA ATGCTTTTG GTTAACTCC TCAATCGAGA AGAAAGTTTA 1500
GAGAGCTCA AAAGTAGG TTCAATCTTA GTTTTCTTC AGAAATATG GTTTCATGTT 1560
25 TCTTATGAC TCTGCTCTT GTGCTTATAT CATCCAAAA GTTTTAAAA AGCTGCGAAA 1620
TTCTATTTTA ACTGATGTT GAGCACTTT AAAAGCTTG TATGTGTGT GAGCTTATTC 1680
TTAACTTTG AGCAATTTG CTGTGTAGG CGATCGCCA CTGTAAAGTT CTTAGATTCG 1740
30 CTTGTGTC TCTGAGATG GAAATTAAC AAATTAAGG GTTCACTCG AGGCTGTAT 1800
TAGCTGTGA ACTATATGTT AATCTGTGT TTAAATTAATA TTATTACTGT GAAAAAATA 1860
35 AAAAAAAAC AT 1872

(2) INFORMATION FOR SEQ ID NO: 23:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 289 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

50 CATTTACCA CATTACACA TGTTCCTT CTCTTTGTT GGTGAGATG AGTGGCTCT 60
TGTCTCTAG TAGAGCAAT CATTACATAT GTGCTTAAA TTCTCTCTT TTTGTCTAG 120
AATATGCTC AAGCTATCT TCTCTCTTT TCTCTACAT GATTTCCCC TCTCTACTAG 180
55 ATCATCTCG TCAATTAAG AACATGTGT TGTTCCTCT AAGAGTACTG TTTCTATATC 240
TAGATATGAC TCTAGTAGA GTTAAAAAA CTTAGGGAG 289

180

(2) INFORMATION FOR SEQ ID NO: 24:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3533 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

10 TTTTATTTAC TTCAATTTTA CTGTACTTTA GTCAATTAAGA AAAACAAATTA TTTTCAATT 60
15 ATGAGCTAC ACAAATCCA AATACATG CTGAGGCTT TTTTATGTC GAAATGTCTA 120
GTATTTTCA AAAAGTAGAG AGTTTACAGA TATACAGCA AATTAAGGCG AATTTATGCA 180
AAGAAACAA GTTTAATTTT ACATTGATG ACACGATTT TTCTGAAAAG CAGATCTTC 240
20 ACTCTTTAA GTTTGCACC AACCGAAT AATTCCAAAC GGTCTTGGCG ATGACCGAGC 300
TGTCACTCT TGTTTATGTC GGAATCGAG GTTATGAGAG CCAAAAAAG TCGTATTAAC 360
25 CTATTTTGG TAGACAAAT TCACAGACA CGACCTGCT TTTAAACTT GCTCTCAATT 420
ATGTACTCC TTCCATCGAG TTGGGAAAA AAAATGCTG GGAATGTCTA GTTAACTAC 480
CAATGTTTC ATTCAGAGG AACTGATAC TCAGAGTAT ACAGTATGCT GTATCCGCTC 540
30 CTTGAAATG GGCACAGC TCCGAGGTT GCGAGAGCTT CGCTGCGCC CTGAGAGGAA 600
GCCGAGACA CGGGGTCTAC GCGCGGCTC AGAGGTAAA GATCTCTTC CGACAGCTT 660
35 CCGCGTGA TAGCTGCCA TCTTGATTC CGGAGACAG AAAATTCATG CGAGAGTAC 720
GTGTGCGCG GTCTTCTG TAGACAGAC CTGAGGTAC AATTCTATT GATTCCTCC 780
40 GATCTGTGA GAAAGACA GAACTCTATG GTGAGACTG ATGAGATAC AATATCTGAG 840
AAAAAACT ACAGGTGAC AAACAGATG TTTGTGCTT CAGCAAGAA GTTTGTAGAG 900
GGGTGACA GTGACTTACA TACGAAAC ATGTACTACA GCGAGTCTC TATGTTCOA 960
45 CATGTCTAG AAAAAGATAT GCTGGATCA CCAATCAAT CAGTCACTT GTTCAATTT 1020
GGCGAAGTT TTATCGGCGA ACAAATGCA CTAGGCTTC CAATGAGGG GATGAGCAC 1080
AATACCTTC AATTATATG CAGCTTTCA CAGGCACTC AATTACGAG CAGCTCAG 1140
50 CGAAGAGAG GGTATCCAG AATGTCACT CAGAGGCTC CATCTCAG CAGGGTATTT 1200
TTGCTATGA ATCTATGAA TATGATTAAC GATCTCAG TTGTGAGGG CATGAAAT 1260
TTGCTATGA ATCTATGAA TATGATTAAC GATCTCAG TTGTGAGGG CATGAAAT 1320
55 CCGAGAGAG GAAATCAAT GAGAGTTCA GGTTTAAGTA GCGCCAGAG AAGCTGCCA 1380
AAGATATAT GTTATCCAA GAGAGAGCT TCTGAGAG CTTTTACTT GAAAGTATG 1440
TCTGATTTG GATTAAGAG GAATCAGCA TTTGATATA ATTACTCTT ATCAAGTAC 1440

181

1500 ATTGTTAATG GAGCAGACG AGGTGAAAT GTGACAGGAT TGGACCTTTC AGATTTCOA
1560 GCATTACAG ACCGAAACAG GAGGAGAGA AGTGTATACC CAACTCCATT AATTAACCC
1620 TTGGCTGGA GAGCTCCTTA TGTGTGAATG GTACAAAC CAGCAATGA ACATCCGAG
1680 GACTTCTCA TACCAATGA AGATTTCOA GCHTACAG GCTCCAGCTA TAAGATCA
1740 ACATCAATG ATGATGAGC TAACTTAAT TTGATACAT CTGGCAGAC AACTTCAGT
1800 ACGATGAC CCAATTCOC TCGAGTAA AGTTCACAA CACAAATTA TAACACAG
1860 AAAAAAGCA TCCAGGTGT ACCTGATGT CGGTACTA ACATTCCTCA AGGATGTG
1920 AGGACCAAT TTGGAATGAT TGGCTGTTA ACATTATCA GGGAGCAGA GACGACCA
1980 GGAATGTAC ATCTTGCAAT AGGAGTGC TTACACAT TACGCTCA TCTGAAGCT
2040 CCTGAATC TCTACCCA ATTGCTCA CCTGGCAT CTTCACCTTG TCGACCTCA
2100 GACATAGACT TCCATGTTC ATCTGATAC TTACGACA TTCACTTAG GGTATAGCT
2160 GCTCAATA AACTTGGCG ATATGTGTA GACTTCTCT TCTATCTCTA TTACATGAT
2220 GGAGGAGC TATTACACT TTTAGCTCA GTGAGCTT TTACGCTCA TTGAGATAC
2280 CACAAAGAG AAGGATATG GATTACAGG GCACAGCA TGGAGCCAC AATGAAC
2340 AATACATG AGAGGGAC ATATACTTC TTGACTCTC TTACTGAG GAAATAGCT
2400 AAGGATTC ATCTGATA TGCANATTA GAAGAGGC CTCACCTGCC ATCCAGCTC
2460 AACTACAC CTGCTAGCA AGCTTCTTA AAAAAA AAAAAA AAAAAA
2520 CCTTTCTT GGGTATGCC TGTCTAGCA CATACTCA CATACTCA GACTGAT
2580 GGTACAGCA CCTGTGTTT AATCTCTCA GATCTGCA ATTGCTTAC GCAAGGTC
2640 ACCATTCAG GTCTGCCCT ACTAATATG TCTGCCCA CAACATAAT TGTATTTGT
2700 TTTTCTTAG TTGAGCAG GTCTGAAT TTCTATAT TTCTTTTT GCGAGCAG
2760 AGACTGAGT GTTAAGAC AAGCAATAC ACTGACAA GTTTACATA GTTTTAAA
2820 TGTAAAGAG AAAACCCCA AAGCTCAA GAATTAAG AACAATTA TGCATGTC
2880 ATTGTAGAC TATTGTAA TAAATACAA ATGTTGTGC ATTTTATGT GAAGTCTT
2940 CTGCTATTC ATTTGAAAG ATGAGCAGA GGTCTGTC CTTCATTTA CTTCCTTC
3000 TGTTTTGA AGGAGTTTC GCGAGTTA ATCGAGAT ATCTGACT TTAGAGAA
3060 GATATTGCA CATCTCTG ATGTTTTC AGGTTGCT TATTACTG CTTCACTT
3120 CCAATAGC CAAACACTG TCCAGCTTT GTTAGATTT TGTATTAAT GTGTACTT
3180 TTTTAAAT TTGAGATC AATGATTA AGTATGAT GTACGATGT GTATATTA
3240 TATATGAC ATCTATTTG GAATGTTT GCTCTGCT ACCTCATTT TAGGAGTGT

60

182

3300 GGTGATGC AATATATGA ATGGAGCAT TCTGAACTG CTGTGAGCG GACTTTGAG
3360 CCTGTGAC TAAAGGACC AGATTTTAC CAGCAGGA CATCATACC CAATGAATG
3420 TGTGGAAT TAAAGAAAT GAATGATCT ATTTACTCT GGTGTTTGA ACAGAGGCT
3480 TTCAATGA GAGAAAAA GATCATCTT GTTATTCTG ACCACATTA GGTCTTCTT
3513 CTTGTATA AATGAGAAA GCTCTCTCA AAAAAAA AAAAAACT GAG
(2) INFORMATION FOR SEQ ID NO: 25:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1148 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:
60 ACCAGGCT CGCAATTA TACTCTCA TTCAATAT GTTGATCA AAGACTTGG
120 GAGCATTC TCCAGGCT TTAAAGAT GAACATGA TTTCATGCA TCCATAGA
180 AACTGTTT AAAAAATT GATCTTAC TTGTCATAC ATGAAAAA CACTGCTAG
240 AATATAGA CTATATGAT CTGTCCAG TCCCATGT CACTCTTGG TCTCATTTCT
300 TCCCTTGT CTTAGTCTAT CCATTAAG CTGAACCA TAAGAGAT TACTTTATG
360 AATATGTT GCATTAAT TAGCATTC TATCTACA AATTAATAT AATTTCCAG
420 ACATGTAA ATGTGTTTA ATACCCCA GACCAATG AATTTCAA AGTCAATAC
480 AGAGATCA TGAAGTAA TTAGTCTA TATTTTCA GTTATTTA AACAAGAA
540 CAATAGTG GAAGGAGC TATTACCT TACTTAGT AACATTCG TTAGTCCCT
600 TTAATACCT CTTATACCA GACTTCCAC CATATATAC AAGTCTTGA CCAATCCTT
660 GGTACTCA GTAAAGTTA CTGTACTAG AATTTTTA TCAATTAAT GCAATTAAT
720 TCTTTTAA AGTAGTTCT TCCATCTTA TTCTGACT GTTCCAAAT GTTCTCCTT
780 TTGAATCA GGTTTTTTG TTTTCTGA TTTCTGAA AATCATCA ACTTTGCT
840 TCTATCTT TTTTGTAT TGTAAAGAT GTCCCTGCG CCAATGGA GAGAAATGT
900 TTAATATG CTTTATGTT TAAATTAAT GAATCTTA TATTAATCA TGTTAACAT
960 TTATGACC TTGTAGTT AAGGTTGA TTTTATAC TTGAGATTT TTTCCCTAA
1020 CTATCTGT TTTTACTT TAAACTAT GGGAAATAT CACTGCTG TCAAGAAAC
1080 GCAATTA TTAGTAGT AATGTAAA GTCCAGTGA CAGGCAAT TTTATATA
1140 TAAATTCGT GTTACTATG TGAATAAA AAAAAAAA AACTGAGCG GGGCCGTA

CCCTATTATA

1148

5 (2) INFORMATION FOR SEQ ID NO: 26:

10 (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15 (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GGGAGAGCT AGCTGCGGACC ACCGGAACAG CTGTGCTGAG TGGCCCGGCT CCTTCCCCC 60
CGCCGAGTCA TGAACCTGCG CCCCTGACTC CTCGCCGCTCC ATCTGCTGCT GGTGCTGTG 120
CTGAGTGCAG CGGTGTGCGG GGTGAGGCT GGGCTGCAAA CCGAAAGTCC CTTCCGGAAC 180
CTGCAAGTGG AGAGCTGCT GGAACCCCCA GAAACATGTC CCGAGCCGCG TCGTTTTCGA 240
GACAGCTTTC AATTAACACTA CAGCGGAGAC TTGTGAGTGG GAGCTATATT TGAAGCTTCC 300
CTGACGAGAG ACCCTTGTGT TATTGAACTT GCGCAAAAGC AGGTGATTCC AGGTCTGAG 360
CAAGCTTTC TCGACATGTC TGTGGGAGAG AAGCGAAGGG CAAATCATCC TTCTGACTTG 420
GGCTATGAAA AAGCGGAGAT TCCACATCT GTCCGAGCGG ATGCAATGCT GCAATATGAC 480
GTGAGACTCA TTGACATGAT CCGAGCGAAC TACTGCTTAA AGCTGTGAAA GGGCATTTTG 540
CCTCTGTGAG GATGCGCAT GGTGCGAGCC CTGCTGCGCC TCAATTGGGTA TCACTTATAC 600
AGAAAGGCA ATAGACCCA AGTCTCCAAA AAGAGCTCA AGAAGAGAA ACGAAACAG 660
AGCAAAAGAA AATTATTAAT AATTATTTT AAAAAAAAAA AAAAAA 717

40 (2) INFORMATION FOR SEQ ID NO: 27:

45 (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1099 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

50 (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GGGAGAGCC GATGTGAGCA TCAATCTGTC TATGCCCATG TTCCCTGCCC TGTACCTAAT 60
CGCCGAGTTC ATGCTGCTGC AGAGGAGCT CTTCACCAAT GGTCTGTCCC GAGAGATGAG 120
GGCCTTAAC AAGATCAACT TCAACACCG CTATTGTATG AAGAGCTCA TGAACATGTC 180
CGCTGAGACT GTGCTGCTG TGTTCAGAT CTCTCTGAG ATCATCTGTC CCGTGAAGCT 240

CGGTGCTGT GAAATGCTTC AATGACAGC CGAGCTTCT GGTCTGATAC TTCTCTCTG 300
GTACGATGAC CAGAGGAGAG TAACTAGTAA CTTTCTGGGT GCAATGTGAG TCAATCTAAT 360
GAGATCTTTC TCAATGTGTT ATGGGAAAT GGTGCCCCAG ACAAATGTC GAAAGAGTGT 420
CTGTGCTTTC ACTGCAATCA TGGGTGAGAG GTGCACTGCC CTGTGATGAG CCGTGTGAGC 480
CGAAAGCTTC GAACTGACCA AAGCGAGAAA GCACTGTCAAT AACTTCATGTA TGAACATCA 540
GCTGACAGAG CGAATCAAGA ATGCTGAGAG CAATGTCCCT CCGAAACAT GATTATCTTA 600
TAAAGACCA AACCTGCTTAA AGAAGTTTGA CCAATCCAAA GTGAGGAAC ACGAGAGAAA 660
GTTCCTCCCA AGCTTTCAGC CAGTTTGAAG AGCTGCCAG ATGGAACAGAG GAAAGCTGTA 720
GTGACGAGC CAGACCTGTC GTGACCTTT CAAAGATGCA GAATGTATG TATGACTTAA 780
TGCAGAACT CAATGACCGG ACGAAGAGC TGGAGAACCA GATTGCGAG CTGAGATGTA 840
AGCTGAGACA TCTGACCGCC AGCTTCACT CCTGCGGCT GCTCATGCC GAGACCTTCC 900
GCGAGAGACA GAGGAGACTC CTGTCTGCA TCAATCGAGG CCGAGGATGTC AACGTGCGAG 960
TGGGACAGAC CGAGACCCCA ATGTGCAATA GCGCCATGAG GATGCACTCC AGCTTCTTCC 1020
GAAACCCGTA CAGAACTTCA AGCAGTGTCT AATTATATCT CCCCATCCA GAAACATTTA 1080
AAAAAAAAA AAAAAAAAAA 1099

30 (2) INFORMATION FOR SEQ ID NO: 28:

35 (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 941 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

40 (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

AATTGCGAG AAGACGAGC GAGGGGTTTC CTGTGCGGAC TGCAGCGGCG GAGAGGAGCC 60
CAGTGAAGCC GCGCTCCCCA AACCCGACTG CCAATGCTGA CCAACGAGCC CTGCGGCTGC 120
TGAATGTATG AATTAAACCA CTGTGACCA TGCACCCGAG GCGAAGAGAG ACCTCAATGT 180
GGGCTATGTC TTGGGGGCGCT TCTTCTCAAT CAGCCTGATC GGGGTGTGAG TGCATGTGCT 240
AATTATGTA CAGAAAGAAA ACGGAGTGA CCGGCTGCGC CATGACCTGC TCCCAATGTA 300
GAGCTATGAC CGAGCTGAGG AACTGCAATG GGTGAGAGAG GAGCTGCTCT CTGACATGAG 360
AGACCCAGAG GTGATGATG GCTTGGAGAG TGGCTAGAG CAGAAAGGCA TGCACACTCT 420
GATGTGAG AGCTGAGCTC ACCCTTTCG CCAACCTTTC AGAGCTGAGG GTTCTGAGCT 480

185

5 GCGTGGGGC CTGCACTGT CTTCCTCTGC TGTCACTGG STGCCCTGC TGGGTCTGG
 GTCTCCATTT CTGCTCTCAC CGACCTCTAG CAGCCTCTGC TTCCATGCC CTCACATCA
 10 CCTCACTGCC CGACGCCCTT CTGCGCTTGG TGGGTGTGA GCTCAGGCC CAGCCACAGG
 CACTCATGG AAGAGCCTTT CCTTCTGGGA TGGGCGCGC TGGTAGACAC CTTTGCTTTC
 TTAGCGCTTC CTGGGCTGGG CTTCGGACA AATCCGAGG CAGGCTTGG AGTTGTTTCC
 15 ATGGTGATGG GGCAGATGT ATAGTATCA GTATATATTT TGTAAATAA ATGTTTGTG
 CTTAAATAAA AAAAAAAA ATCAGAGGG GGGCGCGTAC CCAATTTCC CCTATATCA
 20 ATTGTATTA ACATTTACT TGGGCGCTC CTTTAAANA C
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 756 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 29:
 30 GGCACAGGA AGCTGGAGG GGGCGGGGT CGAGTCAGG GGGAGCGAG CTCTCTGGC
 TTGGCAAGG GGGACTGGC CTGCGAGGG ACCGAGCCA CAGACACT GGTCAAGGA
 35 GTAAGCAGG GATTAACAC TGGAGGAGA GCGAGGAGA AGTCACTG GCTTCAGGT
 CTGCTCTGG AATCCAGAT AAGATGCC ATTTTCCAC ACCAGGAG CAGAGCCTGT
 TGTTTGTTC AATATCAA CTCACATCC ACAGAGAGA GATCTAAG ATTATGCGG
 40 AATCTCAGA AGAAGTTTC TGGAGAGAG CTCTGCTTT TTCTTTGTA AGCATGCTG
 TCACCCAGG ACTAGTCTAC CAGGTATTT TGGCAGCTAA TTCTAGATTT GATCATTC
 45 CCAAGTTGC ACTTGTGT CTCTGGAT TTGGCTTGG AAGGTATCA TACATPAGG
 TATCCAGAG TAATTTCCAT TTTTTCAG ATCAGCTCG TGGGCTGTG TTTGTTCAC
 50 AGCATACAG GCATGCTTC CTTACTGTG AGCATGCA AATTAAGCAT GATTAAGTG
 AGAAGGAGA CTCTAGCT TCAGCTTCT AATTTGTG TCTGTACTT TCGAATTTT
 55 TTAACTCT GAATTTGAC ACATTTAAA TTTCAAGTGT ACTTTAAAT AATATCTC
 TAATGGAAA AAAAAAAA AAAAAAAA ACTCGA
 (2) INFORMATION FOR SEQ ID NO: 30:

186

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2100 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 30:
 60 NCGAGGCA GAAGTCTGT CTCTGGGG GTAACTACA GGTATCTTT GGNACAGAG
 120 ATCTATTTT GGAATTTACT TCCAATGATG CTGTGAGATT TTATCCCTGG ACCATTGATA
 180 ATAAATCTA TTCAAGCAG ATCAATCTAT GTGTGTGTC AACAAATTT CTGTATCTG
 240 CAGAGTTGC AGAATCTGTC CAGCATTTG TGTTTTACTT TCAACGACA CAAAATCGG
 300 GCGTTGATG TGTCTCTCA TGGCTTCAC TGGCAAGGC ATGTTTACC GAGGTGATGA
 360 TCTTGTCTG CGATAGAGTG TCTGAAGATG GTATAAACCG ACAAAGCT CAAGAATGCT
 420 GCATCCAAAC ATGGCTTTGA ATTGTTAGAA CTTAGTCCAG AGGATTTGCC TGAAGAGAT
 480 GATGACTTC CAGATCTAC AGGATTAAG CGAATTTGTC AAGCCCTGAA TGGCANTGT
 540 TGTGTCATG TAGTGATGA GAATGATGG AACCAAGCT TTAGCTTACT CCAACTCAT
 600 GACTGACA AACCATGCA TTGGTCAAG AGATCCCTGT CAGCCAGAG AACCCATTT
 660 GCGAGGCA GATGATCTG AATCCCTCT TCAATCTCG GGTGTGCAAT CTACACAC
 720 AGATCCGAG GTTGATGCA TTGTGATCC CATGTTAGAT CTGATATTC AAGATTAAGC
 780 CAGTTTACC ACTGAGGAG GAGATGTGA GAATTTTGA AGACTCTTT CAAGTTAAA
 840 GGAATGAA GACAGGCTG CGAGCTTCC TCAATGCA AGAATGTC ATCGAGAAA
 900 GTGGCCAAA GCTTCTGGA TGGCAATCG GCGACAGA GATGAATTC AAGGCTTTC
 960 ATCTGATGA GAGCACTGA TTATTCATAC TAGGTTTGA CCAACAAAG TCGTAGCTGT
 1020 CTCTGATTA CCTCTACT CAGCCAGTC ATATTTTGC AAATTTGCC TTATCATGTT
 1080 GGTGCTGTA CTGTGTTAA GGTGCTCTT AATTTTATG TTTTATAGA GTTTAAGGAG
 1140 AATCTTTTT TTCTCTCAT ATATTTTAG AGATGAGA ATACAGTAT AGTATAGT
 1200 GAGATTTCT TAATTTACT TTTTTTTT TCTAGAGTG AGGTTAGAT AATCTCAGA
 1260 GGTGTGTG ATTTACTGA GTTGAAGA ACCTCAGGC CATCTCTGT CAGCTTTTA
 1320 AGTAGATTT CAGCTTCA CACTGATAC TGCACATCG GAGTGTGTC ACCTTTCTG
 1380 GTGTATTTG GTTTTCTCA TTCAAGGAG TTGTAGTCT GAGCTATGA TGTCTTTAT
 1440 GGGAGGAG GAGGAGCTG CAGATTTAT GTGAGCTAG TGGGGCGAA GTCTCAGCC
 1500 GCGATAGT CTCTACTTA GAATTCCT CTGGGATTC TTTTGAGTA TAGTGTCTGA
 1560 GCTCATCTA GAAGATCA AAGAGCAAT GTGATTTTT AGACTTAAT AATGAGCA

(2) INFORMATION FOR SEQ ID NO: 30:

187

AAAGATTCT ATTCCAGTGG GAGGAAACC TCTCTACTGA GTTGTGGGGG ATATCTGTGA 1620
TGTTAAAGAG AACCTTAAAG AGTCTGTGA TGGGGCAATG AGACATGATG TGAATACATA 1680
CCGTGATTTT CATAAGAAA TTCTCTGTC TTAGAGTCTT CCCCCTGTCG TTGAGATGCC 1740
AGAGCTGTGT TGTTCACAC CTCGAAAAA AGGACATTT CCCCCTTCTT CTTTAAAGCC 1800
AAAGAGATAT CACTGCCAAA GTGGAGACAC TAAAGGGTGG GTGGGAAAGT GAATATTTAG 1860
GGATTAATTT CCGAGGACC TTGTTTTCT TCGAAGTTC GTAGCTCTTC TCTGCCCTTC 1920
CAAGCTGTA ACTCGGAGAG ACTATCTTTT GTTCTTATC CTTTGCTTG TTGGAGTGGG 1980
TCAACCCCGG AGAGACTGAT AACGAATGG CAAATTTTGA AAGGAAAGAT GAAAGATACT 2040
GGAAATAAA ATCTTATTTT GTTTTGTGAG AAAAAAAAAA AAAAAAAAAA 2100

(2) INFORMATION FOR SEQ ID NO: 31:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1448 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

AAAAAAAAA AAAGCCGACC TGAAGGCTGG TCTCTTCCA GTTGTGTGAC CCTTCAGCTG 60
GGATTATCCA GCAATGATTT TTTTCAATAGT GCAATTTTCC TTATTTTCAAG GATTGACCTT 120
GAGTGTGATTT TTTTGTTTT TTAATTTGTT TTGTTTAAA ATTAATTTAAA GACAGTCCAG 180
AGCTTTTCAAG CCAATTTGTC TCTACTCTGG TGTAAATAT TTTCCTCTCG GCGAGGGGAG 240
CGAGGTATGA GGAAGAGAG CAAACAGAG TCGAAGGTGA GCGCTTCTCC TCGTGTACT 300
AAACAGAGAG GTTAAAGTTC CAGCTTAAAG GATTGTAGAC CCGTTGAGGT TCAAGGAACT 360
GCTTGGCCAG GATGAGATGT GATGTGATAG GCGCAAGGGG GGAAGAGGGA AGGTGAGCCC 420
CGAGCTCTCC GAGATCCGAC TGGATCTGAC TTAGAGGGGG GTGGAAAGCC TGTCTGACC 480
GTCTGAGGGG TTGTGAGCCC GCGCCCTCTC CTATATACAC CCGTGGAGCC AGGAGATCCC 540
AGAGAGAGAG AGCGAGAGAG GAGTCAATGG GAGCCGAGCC TCGAGTTTGA GCAAGTTTCA 600
CTATCTCAAT GCTGGGGGTAAC AGCAGAGAG TACTACTATT TCACTGTCTT TACTCTTAA 660
TTGGGGCAATG GCTTTCATCC TGTGTGCTCTT GACTGTCCA GATGAGTGGT AGCGAGACAC 720
TGGGAAAGTGG GAGTGTGCTT TGTGCTCTCC TTCCAGATGG GCTGTGTGTA CTGCTGCTCC 780
CGACCCCTTAC CAGTGTGCTCC AGGAGAGAG GAGAGTGGGG GAAAGCAAGA TTGGAAAGAC 840
AGGAAAGCA AGGCTGTGAG AGAGCTCTCT GTCTTCTCTC CACTGTGATG CCCCCTGCTT 900

188

GATGTGCTGG TAAATCTTTT CTCACACCAA ACCCTCTGCC AGGACAAAA CAAAGCTGCC 960
TCCCTCTCTT CCGGAGACTG GTGAGAGCTT TGGGCTTTTC AGTCCCAAGG CCGCGAATGG 1020
GAGTCTCCCT CCGACTCCAG ATTAAGACAG GCGCCAGGCC TGAAGGTTT GCTGTGCCAG 1080
GAGCGGCAAG CTCTCTGAGG CAGAGCTGTC CCCCCTCTTC CTCACATCTT CTTGATCTCG 1140
CTTCTCTTTT CTTCCAGAT ATTAAGAGA ATCTGGCATT CTAGCTGAG ACATTTATAT 1200
TGTTTTATTT TGAATATGAT GTAAATCATG AAAGCTGTCT GAACTATGAT TTGTGTGAT 1260
ATATTTAAA AAAAAATCAG TGTTTAAATA AAGCATATAG TACTATATCC TTATACATCG 1320
CGATAGCAT TTGGTATGTA GTGATTAATCT GTGAATATA AATACAAAT GAAATCTTAA 1380
AAAAAAAAA AAAAAAAAAA AAACCCCGGG GCGGCCCCCG GCGCCCAATT 1440
CCCCCA 1448

(2) INFORMATION FOR SEQ ID NO: 32:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 456 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GGGACAGCA ACTTGAGGCC ATGAAGATCC GGGTCTTCC TGCAGTGGTG CTCTCTTCCC 60
TCTGTGCTCT CCACTGTGCC CAGGAGACCA CCGTGGGTGG TCTTGAGAGA GAAAGCAGCA 120
TTGAAATTA TGCCTACAGA CCGAGGCT TTACACCCC GTTCCGAGAC ATGCAGAAAT 180
TGCATCTCC GTTTAAGCT GATGAGTCC TGAATGTGCA GCGCCCTTT GATCTATCA 240
AAAGGAACT TCTTTCTTC AACTGGGATG CTTTCTTAA GCTGAAGGA CTGAGAGAG 300
CAATCTGTA TGCCTAGTGA CAGTACTTC CACTGAAAG GGGGGCTTAC GTAGAGGCTG 360
ATTCTAAC TACATTAAT CTTTCTTCCC TGAAGAACT CATTAAAAA TTTTCAATCC 420
AAAAAAAAA AAAAAAAAAA CCGGCGGGG GCGCCG 456

(2) INFORMATION FOR SEQ ID NO: 33:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1176 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55
60

189

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

5 GGCACGAGTG CAGGCCGAGA GAGGACTCAT TCAGAGGAGT GAAGGGGAG GTGGGTTTTT 60
CTTCTACCC AACTTATCCC CTGTGAGCTG GACAGCTTGG TAGCAGCTCC CTGAGCTTAG 120
ATGGTGTAG CCAAGAGAGC TCAGTTTTTA GGGACAGGA CGGGGAGAG AAGGCTCTGG 180
CACACACAGA TGTGTCCATA TGTCTCCGAA TGGTCTGGG ACTATTGCTA GCTAGAGAC 240
CCTAAGTCTC TTCTTCTCCA TGTCTMTCT CCCCTGTSTC ATGGGGCCCTA AGTCTCTTTT 300
CACTGGGCTT GCTTCATGTA AGTGTCTGCC CAGCTAGCCC GAACACAGCC AACTGCGGCC 360
TATCAATGCC CGAGCTGCA TGGGCCATCT TCCGCCAACC AACCTGGCTG GGGCCGTGG 420
CTCCGACTG AGAARAAAS TTGGCART CAACTGGGCC CGGGCAGGAC TGGGCTTCCC 480
TCTGATCAT GAGGTGGTG ARCCAGAGC CGGAGCCCC CTACAGCTCT GACTTCTCTG 540
ACTGTCTAG TTTTATATGCC AGCAGTADCC CTGGACAGGA GGAAGTAGAC AGCGCTCTG 600
CTGCCCCAGC CTCTACAGC CGAGCCCCCC GGGCCCCAGC TTCCCCAGGC CGGCCCCAGC 660
AGCACACAGT GATCCACATG GGCATCTCTG AGCCCTTGAC TCAGGCCCTT AGGAGGTGT 720
ATGATAGCG GATGATGAC GGGACACAG GCTTCATGAG AGACTGTGAC GATGACAGT 780
ACCGAGCTG GCGGGCTTG GTTTGGCTGG CCGGCTGCT AAGGAGCGCG GCTGGCTCTC 840
GGAGCGRCC GCTGACCTTG CTTCAGGGGG CGGGGCTGCT GCTACTCTTG GACTCTCTGG 900
GCTTCTGCG CTTCTTTCGC CTGATGTCT GCTTAGGGCG GCGCGAGCT GACAGCGATC 960
CCAGCTGGA CCGACTCATG AACCTTACA TCGGCTGGG CCGCTCTGCA GCGCCCTTGC 1020
TTGTGGCTAG GCGAGCTAG GATGTGGTT CTGTGAGGA GAGCGGGGT AATGGGGAG 1080
CTGAGGGGAC CTCTTCACTG CCCCCTCTCC TCAGCTTAA GACACTHGA CCCCAGCC 1140
AAGCCAGT CCACAGAGT GCTGACGCG CAGGCTGGA GTCCCGCTGG GTCAAGCATT 1200
TGTCTTAGT TGTCTTCTCT CGGGTTTCC AGCTCTGAC CCCCCTCCAC ATGAGAGAC 1260
TGGAGGTGG AATTAACAA CACTTTATT AAAAAAAAA AAAAAAAAAA 1320
AAAAA 1326

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 710 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

190

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

5 GCGAAGAGA AAGAGCTGG AGCTCCGCC CCGGGGCTG TCAGATGGCT TGGTTTTCTG 60
CGAGCCGATT GGCCTGGGGA GGGGAAAT TACTACGAA ACATGACTAT TTTAGCTGC 120
TTAGCAACAG CTCACCAAG TAGAGAGAC ACCAGGTAG GCAACCCAGT GTGTGCATCC 180
TGGGCTTGG GGCAGCTCT GAGAGGCA ACCTCTGCG ATGCAATACT TCCATTAAAG 240
AATGCTGCC CTCCTTCTC TCTTATCTT TTTCTTTTCA ACAGTGTCTT CTTTTTGTGG 300
GATGCTTTG CGGCAACAA CGGGCGGCA GGCACACAA GACACATTTG CCTGCGGTA 360
GACAGGGGG GAATGTGAT ATTTTATTAA GCGCTTAAC AATTCTGAA ATTCTCAAA 420
GAAAGCCTT TCAGAGGAC CTGGGCTCA AGCTGCAACA AATACTGGA GTTCGGCTC 480
GCATTCGAG GCTGCACCA ATAAAGACAG CGTCTGGAT ARTGGGCCAG TGTGTGCCAG 540
ATTTTTTTTT CTTCTCTCTT TTTCTTTTAT AACTTAAGGG AGACTTAGG CTCTTCCAGG 600
GACACGCC TCGCTTAG ATAAACAGAA TGGAAATTA AAGAGGAAG CAGGACGTT 660
GGGAAGAGC ATCTTCTTAA AATCTGCTT GCGCCGACG CGCTTTCTCC 710

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1188 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

40 GATGGCTTTT ATATCTTATTA TCGACCCACA GACAGTGACA ATGATAGTGA CTACAGAG 60
GATATGGTGG AAGGGGACAA GTACTGGGAC TCCATAGCC ACCTGCAGCC AGAGACTCC 120
TAGCAATTA AGATGGAGTG CTTCATGAA GAGGGGAGA GCGAGTTGAG CAGCTGTAG 180
ATCTGTAGA CCAAGCTCG GAGTCTTCT GGCAGGCTG GTGAGTACC ACCCCCACT 240
CTGGCCCCAC CACAGCGGCC CCTTCTGAA ACCATGAGC GCGCGGTGG CACTGGGGCC 300
ATGGTGGCTC GCTCAGGGA CGTGGCTTAT CTGATGTGG GGTGTGCTCT GGGTCTCATC 360
GTTCATCA TCCTACCTT CATGCCCTTC TCTTGTGGA GGGCTGTGTC TAAAGAAAA 420
CATACAGAG ACTTGGTTTT TCTTGAAGT GCGTTTCCAC CTTCTGCCC GTATACTATG 480
GTGCATGAG GAGGACTCCC AGGCAACAG GAGTGTGACA GCGCTTACTC AGTGGCATCA 540
GTGAGCGGCC CTGTCTTAT GGAATCACA TGAATAGGG GTGCCCCCTG GCTCAGTGG 600
GCTACCCGGG CATGAGCCC CAGAGCACT GCGCAGGGA GCTTTCAGAG CAGAGTGACA 660

191

CGACGACCTT GCTGAGGACG ACCGATCTTG GGAATGAGTA TGAACCCGAA AGTCAACGAA 720
TCAACAGAGG TCCGAAAGCT AGCCGAGGAG AGGAGCTCTT CTATACAGCA CTGACCGAGG 780
5 AGTCAACGAA CGACGACCTG CAGCCCGATC AGGACCTCTG CGAACCGCAG GAGCAGCCTG 840
CTGATCTGAG CGATCTCAGG GTGAGAGAGG CCCCACAGAG TCTCTCTCTG GAAAGCAGCT 900
10 GGAACCTTC ATTCACTCA GGGCCCCAT GCTGCTTGGG CTTGTGACCA GTTGAGAGAG 960
TGAACAGCTC TGACTCTGCG CAAGTGAAGT GAGAGAGCTG GTGTCCCGAG CAGCCCGTAG 1020
GGGACCTAGT AGGACAGGAA CTTGGAATGC AGCTCTCCCC GGGGCGACTG GTGCTGTCT 1080
15 CTTTGAAAC AGCAGCTCTC AGCAATTGAG CAGAGCTGA TATCCGAGAA AGACTATATA 1140
TTGTTTTT TTGAAAAA AAAAAAAAAA ATCTCGAGG GGGCCCC 1188

(2) INFORMATION FOR SEQ ID NO: 36:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 956 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GGCAGACGAG TGAATATGCA TCCATAAAT TCAATGTTTA TACAGGCTC ATGACATATA 60
GATGTGACAT CTGACAGACA GGGGTGACC AGGTGATAC ATCCCTCCCA GATGTGACAT 120
35 CGAGATACA CTCTGCTAGC AGAATGGGG CCCCATCCCT TACATATCTG CTCTCTCTCA 180
AAGTCCAGCC CAGAGAGAC CAGGCTTTTG ATGACATAG GGTGGGCTTC CGACTACTTT 240
40 AGTGAATAG GAGCATGCT GGAATCGAGG TTCTGTGCT TCTGTGAGAA AGGAGATCC 300
ATTGACCCCT GGCAGACAGG TCGAGTATTC CATCTTCTCT TCTGTCCAG CTTATGCCCC 360
45 TCCGACTAG GGCACCCCC ACAACTTCTC CTGAGGAGAG GTTATCCCC AGCTGAGAGG 420
CTTGACAGAA CGACAGTCA CAGAAATCAT TCTTCTCTCT GTACTGGGC TTAACTCCCT 480
GCAAATCTCC GAGCAGTACT GCAATGAGATG CGAGAGCCAC CGAAGATATA CAGAGCCAG 540
50 TTATATATTA ATAAAAGGAA AATATCTAG CTGAGAACT CTGCTTTTGA CCGACCATG 600
GGCAGATCA CATCTTCAAG GCTGTGTGAG CAGCTTCTGA AAAGCAGGC TCGTATATGA 660
CTCCAGACAA TTGCTATGAA GTGAGGAAAA CTGCGGTGAG TCCGAGAGAA TCTGAGGTCC 720
AGGCGAGGGA GAGAGATCA TGAAGATGA TAACTAAAC TGTGTATAT CAGCGGAGAG 780
60 GGTCTTATGT TATGAGTGA AATGAGGCC AGTAAATTAG TTGATCTCT CAGAGATATA 840

192

AGCCTATATA CAGCCCGTAG ATACCGACCA CTGTGTGCTT GCCCCGCTT TCCGATCCA 900
AGATGTCTCT TCTGTCCAG AGACTATATA GGGGCTTCTT GCAATTACTT AAAAA 956

(2) INFORMATION FOR SEQ ID NO: 37:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1603 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

TGACCCAGC CTGCGCTCT GCGAGATC TGGTCTTCT GTAGACCCA GTGAGAGAA 60
20 AGCATTTGAG GATTTAATC GAAATTAGA TGCATTAAAT GTGAGATTC TGAGACTTCC 120
TCTGAAATG GCAATTTCAA AGGAGAAAT GCTTATACC CAAAAGCTC CTTTGAAAT 180
TAACTTACAC AACTGTAT TTATAGCT TTATATGGGT GAAACAGAA TCAAGAGAT 240
25 GAAATTAAT ATGTAATAA TTTCAGCCA TTTCCTCTTA ATGATCAAA AACAATTAG 300
GTTTGATAG GAACTCTGT GTACTTAGA CAGGATAG AGAATCTAC ATATGTCTAC 360
30 GTACTGAG CAAACAGAG GACTGATAC TGTGAGATC TTACAGGGA TGTCTGTGCC 420
CTCTCGAGGC TGTCTGTGGA GTCCCCCTTC AGTGTCAAT TCTCAGCAG TTGTGTGAGG 480
CTGCGACTT TAAATTAAT CAAAGCTG ATTGACAGA GCGAGGTAC TGAAGTTTGG 540
35 AACGAGAG ATGATTAAC TATGAGATG GACCTGTGTA AAAATGCTG GTATCACTCT 600
AATTTGCTT GCGCATCT TGTACAGAA ACAACAGTA ACAATCCAC CATGAATAG 660
40 GTCTGTGTC AATTTATATC AAGAGATGCC CTGAATATA TGTTTATAG TACCAATTA 720
AAATGCTCT ACTGTCAAT GAAACAAAT CCGAGATAG CCAACAGAT ATTTTCTGA 780
45 AAGATTAAT TTATGTGTA ATTTGTAAT GAACTGAAT CTGTGATGCA TTTCAGAGAA 840
GAACTTAC TATATGAG GTACAGAGG ACTGAGGGA GTGTCCCGG TGAATATAT 900
50 GAACTTGC CAACTGTGA GTGTACAGC ACTGAGGGA GTGTCCCGG TGAATATAT 960
CATAGGCTT TATATATAT TTGTGTCTCA TTTCATATCA AGTAAATACA CGACAGTTG 1020
TCAATCAAT CAGCTTTTGG TACTATATA TATGTGAT TTTTTACTTT TTAAAGACAG 1080
15 AAACGAAAT TGACCTCCC GCGATGTGT TAAATCTCT CTGCTTTTGA CTTTGTGCTAT 1140
TTTCTTGA TA ATGTAAAGC TTGAGATGT TTGAGAAAA GTTTATATC CTGTATATGA 1200
TACATATTA AATGAAAT TTTCAGAAAA AGTTGATTA ATGAAATGT GGTATGAGAA 1260
60 CTAAATGCA TTTTATTTG CTGAGAGAG AAGCTGTGA TGAATTCAG ATATGCTTT 1320

193

194

15 5 TGAATGTTTC CTCCTGCA GCTCCAGAA GTACAGACAC CTGCATTTTA GCTCTGCATG 1380
CAGCCGCCAGC AGGCTGCGTG TTTAAGAAAT TCAATGTTTA ACTGCGCTGCT GTAGAGAGTC 1440
TTCCGTTAGC ATAGAGTGA AGAGTACTA TTCTTTGGTT GGTGTTTGT TTGTTTGT 1500
TTTGTTTGT CTTTATGTC CAGAGGTGC TTGTTTAAA AGTATGTTA ATAAATGAA 1560
1603
10 ATTCTAAGT TAARAAGT TCTTAAGTT GATATTAACT TCT
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1089 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:
15 (2) INFORMATION FOR SEQ ID NO: 38:
AGCAGAGCT ACCTTCTGC CTGCTTCT GCTCCACAA GCACGAATC CACGGCTGT 60
GGCTGCTCA CCACGCTCC TCTGGAAC GCACCGCTG TTCTTGAAA ATGCCCGAGT 120
CCTGGTGC AAGAGGCTT CTTGCTTTC CTCTGTGCA TGTGTATCT CAGCTGTATC 180
GGTGCATGG CAGACACCC TCAGTCATCA TCTCTATCAG GACAGTCAG CCTGAACTCA 240
AGTCTTACG TTGCGAGTT CTTTTCTCC TCTTCTGTT GTTGGCTTC ATCCCTGCAC 300
CCCTCATCTT CGGGCTGCG ATCGACTCA CTTGCTGTT CTGGAGGAG TTCTGTGGG 360
AGCAGGGCC CTGGTCTTC TACGCAATG TGCTCTACG ATACTGTAT GTACAGCATG 420
CCATCGGCT CAATCTCTT CCTTCATCC TGTACACAC CAGTGGCAG TCTGAGGAA 480
AAACTATAA CCTACATCA AAAACCAAA GGGCGGGCTG AGCACCAGT AGTCTTTCC 540
CTCTACTCT ACCCTAGACA ACCTGGGAG GGAACCTGTG CCGCAAAAC AGACACATAG 600
45 GACAAAGTTT ATCTATACC TCGAAGCCA TGAGTGTGT GAAACATGG AGTCCGTTT 660
ATAGTGACTA AAGGAGGCT GAATCTGTA TTAGTAATC AAGGTGAT TTCTTCTTAA 720
AAAAAGAAA AAGGTTTCA AAAAAACA AAATCTAGTA CACACACCA GGCACAGT 780
50 CACACACAG CAGACAGACA CACCGATT GTCTTTTTC TCAGCATCAG AGCCAGACAG 840
GATTCAGAT AAGGAGAAA TGACATCTG CGGCAGGCTC CTGGAGGCCA CTGCGGGCC 900
55 TGGGCCAG AGTCTACTT GAAGGACT CATGGTTTC AGGATGCTA CAGCTGCMAG 960
CAACAGGAC TCCAAATTC AGGAACT GTTGGCCAGC TTGGAGGAGT GACTTTCTG 1020
60 GATACACATA CACTACAAAACAGAAACA TTTTAAAA GAGTTTCTT AAATAAAAA 1080

AAAAAAA 1089
(2) INFORMATION FOR SEQ ID NO: 39:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 629 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:
15 AGCTCAGTTC CTTAGAAAT GAATTTTAA ATGCACACTAC CAGGTAAAGC ACTGAGACCA 60
GTGAGGTCA TAGTAGAAA CATAGGAAT TAAGAAATTT TAATGAGAA AGGAGGTAT 120
20 GAATCCAGT TACATCTTAA GACTCACTGT AGTGTGAGT GTTGTAAATT ATCTGCTAT 180
CCATCTCTT TTAATTTTT CTTAGAAAG TCTCTATG GTACCTTGA GGGACTGCTG 240
TCAAATATA TGGAAAGTG GGTCTGTGT GTACAGAGG TGGACTTTCC CACACATGA 300
25 AGTTGCTCC CAGATCTTC ACTAATGAA GAATCAGCA GTGAGCTCA CAGATTAGCC 360
AAATACTGAG CTCATTGAA CTACTAAGC CTGACATTT CTGCTAATC CAGGACTCT 420
30 GTAAATACA GTCTTCTCT TGGAGCTTCC CATGTGTAG CTGAATATT GTCAATCTG 480
CATTAATAC TAAGCTCCA CATCTTAT CTGTCTTTC CCCCTTTTC TTTCCTTTC 540
CCAGGCTCA GCTCTGCTG ATAGTCTGA GACTTTCTT GCCAATCTT GATAAATTC 600
35 TTGCACTCT AACCCATCT CAGTGTCTG 629
(2) INFORMATION FOR SEQ ID NO: 40:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1964 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:
50 AAGAGACAT GGAATTTCT GAAGATGTT TCAGGCATAT TAAGAAATC TTTAGCGAC 60
TTGAGAAAT CAGAGCTCT GAATGCTTC GAATGGACT GGAAGATCT AAATACCTTT 120
55 TACTGAAAG AGCCAAATTT ATTGCTATCA CCTGTACTCA TCTTCTCTTA AAAGCATG 180
ACTGTGCA CTTAGGTTT AAGTATGACA ACATTTGAT GGAAGAGCT GCTCAGATTC 240
TGGAGTAA AACTTTTAT CCTCTCTTC TACAGATTC TCAGATAGA TTTAGCGAC 300

195

360 TAAAGGATG GATTATGATT GGGATGATG ACCATGATC TCCATGATT AAAGGATG
420 GCGTTTAAA AGTATGAAA CATGAGGAG TGTCTGTCA CTGCGTTGT TGGGTTGGA
480 GTTCCAGTG TTGACCTTGA TGTCTAAGG AGAGCGAGG GAGCTTGT GAGTTTCTAC
540 AACTGGCAT ACAGAAATCT AGGAACTTA CCCCATGTC AGCTGTCC AGATTTAGT
600 AGAGAAATG CTGGCTTACT GTATGATTC GAGCTGATTA ATGTGATGA TTTTCAGGA
660 GTGGAGAAAT CTGAACTTAA TCCCTACTTC TATGAGATC TTGAGAGAGC AGAATATGA
720 GTAGCACTTT TTATATGATG GTGTTTACTT GTTTACCTTG CTGACAAAT GATATTTCTA
780 ACAGCATTA ATGGCGAAA GCACTTATTT GCGAGATCA TCAATAGAG ATGTGAAAC
840 AATCCATTA TTGAGAGAC AAAGAGTGG ACACTGTGG ATGATTTCA AGGTCAAGC
900 AATGATTA TTCTCTTTC TGTGTAGCA ACAGAGGAG TGGGCAATCT GAGGATGTC
960 CTGTGCTGG TATGAGCAT GTCTAGAGC AGACTGTAC TTTATATCT CCGGAGATTA
1020 TCCCTCTTC AAATCTTTT TGAATGACT CCAAGTTTCA GTCACTGAC AGCTGCCCC
1080 CTTCATTTTC ATATATTTCC AACAGAACT TTCCCACTTA CTAGAAAGAA TTGAGAGAGA
1140 GCACTGTAG AATGAAAT AATTAATAAT ATGCCGAGA TGGCAACTT TGTATGAC
1200 ATGTATATG AATTGATTA GACTAGATC CATTATATC AGACTTATT ACACTTACA
1260 CTTCTATAG TTGAGAGAG TGAAGAGTT GAAATGAG AAACAGAA TT GGAAGAGAA
1320 GAGAGAGCA TGAATTTCA AGCTGATAT ATACCAATG CAACAGAC GAGTGGCGT
1380 GAGAAATCT GAGCTTTCA AACTGATCC ACCCCAGTG AGAGAGAGC GACTTCCACT
1440 CAGAGAGCA TCCCTGCTTT ATCTGAGAC ACCCTTACT TGGTAGAGC TGTATCTGCA
1500 CCGGAGAGG GTAGAGATC TCAAGATGCC AGATCTGCC CAGAGAGAC GAGTTAGCCA
1560 AACTGTATG CTTCAGAGG AGAGATGTC AGTCAAAA TGTAGATTA GCTGTTTTT
1620 GTATTTTATA TTGCTCTAT CCAATTTACT GTACATTAAT AATGTTAGT TCTTAAATTT
1680 GTTAACTAT TTGCTGTCT TGAATATTT TTTTAAAT ATGTATGTA ACAATCTAG
1740 TTTCAATTT TGAATGAAA GAGAAATTA CCAATCTCT GATGTTTGA TCACTGAGTG
1800 TGTGTATAT CATACCTTCA TTAAATCAT TTTCATGAA TATATATAT AATCTTACA
1860 TTTTATGTA ACTCTTAA AGAGAGAC AGAATTAAT GAGCTTAC ACAGATGCC
1920 TTGATGTCC AATTTGAAA GAACTGTGT TCTCTGTTA TACTATCAA TCTTAAAT
1964 CAGTTTCTTT TTTCATAC TTGTGTTCT GTCTTATGT AAG

60

(2) INFORMATION FOR SEQ ID NO: 41:

196

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1522 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(KL) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

60 CTTGTCCGCG CCGCTGGAG AGCTGCTTC GCGCGGAGG CCGCGCGCTG CCGCGCGCTG
120 GAGGTTGTC GCGACTGGA CACTGTGAC CAGAGATGAG TCGAGATGTC CCGCTGCTCC
180 AGCGCATGGA CTGTAGATTC AACAGCGTT CCGACTGAG GATGATGTC ATCAGAAC
240 TACTGTGTT TGGCTTCTC CAAGCTGTT CTGACAGAG CTTCGCGAGA GAGCTGAGAG
300 GACTGGGCA GAGCTGCCA GTCTGTGTC CCGAGTGGG GGGCTTACAT GAGCTGCAAG
360 CTGATGGAAA CCGAGAGAGC GAGTCCGCT TTGAGAAAT AGAGAGAGAT TCTGAAATCT
420 AAGAGAAAT CATCCGAAAT ATTGCAAGC AGCTGCCCC GATCCGAGAC AGCATGAGAC
480 GTAGCATGCC TCCGAGCTTG GTGAAAGGCC TGGCGCTTCA GCTAGAGAC ACCAGCGGCT
540 CCGAGAGAGA CCGAAAGAG GAGTGTGCA CTGCGCTTGA GAGCTGTCT GAGGCTTAC
600 CTAGAGATC GAGAGAGAG AAGACATGTC TGTGTGTGTC CTTGTGTCT GCGAGAGAG
660 TGGCGATGCA CAGCGCTGCC TTGCTGCTTG ATGTCTTCA CAGAGATGT AATTTATTTA
720 ACCAGAACT AGCGAGTAC GTGAGAACT TAGCGAGAAA TGGATGTGAC TGAACGAGCA
780 GTTTCAGAG TGTGATGTC TAAAGTCTA TGTGTGTCA GCTGTATAC TGTCTTCAAT
840 GTTGAAGAG CCGTGGATG TGAAGATCT TGTAGAGAA GAGAGCTTG AAGATAGCTG
900 GTAGCTTAT TTAAAGACA ATGTTAACT TTAAACCC TTAAATATAT CTAGATTAAT
960 ATACTGAAAT GAAATGTCC AATTAGACT AATTGAAAG CCGTCAATATC ATCCAGCAAT
1020 GATCTGTAC ATCTTAAAT CTACAGAGG TCCCTTTAT TCCAGTGTG AGGTTCCAC
1080 TTAAATATTA AATTTGAGG GAGTTTCA GGAATGAG AACAAATTA AATTTTGTG
1140 TAAATATTA TTACTGTTTA TTAAATTCA ACCATGAGG ATGTCTTAC AGAGGAGAT
1200 ATCTGTTG AGAAAGCA GATGAGAG AGAAGATAC CCGTTACAA TGAATATATC
1260 CACTGAAAT AAGGATATAC TATCTTATTA CTAGAGAGC AAGATCTTC CTGTTTAAAA
1320 AACAGATG CAGAGAAAT CAGAACTAA GTGTGACTT CAATATGAG GAGTGGCCCG
1380 CCGGATGCA CCGAGGCTT TGAATTTTT TGAATCTCT TTCTTTTTT TTGAGATCA
1440 AAAGAGAC TTGTGTGTC TTAAAGGAT GTGAGAGTGA TTAAAGGAG CAGGTGTGAC
1500 CAGGTGAG TTACAAAT CAGGTTACT AACTGTATC AACTTTTTT CATTTGATG
1522 KCATCAGAG TTGTGAAAG AC

197

5 (2) INFORMATION FOR SEQ ID NO: 42:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 875 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

15 TGGGATTTCC CTTTATCATG GAGGCTTGT CCACCTTCT CTATGTCCT TTCTTTGTTG 60
TCTGTGCTG TGGGGCCATC TACTGTGCT TGTTCCTTCC TGAGACCAAA GGCAGAGCT 120
TCCAAGAGAT CTCGAGAGAA TTACACAGAC TCACCTTCCC CAGGCGGGCC CAGGGCCCCA 180
CGTGGAGAG CCTGGAGGTT ATCCAGTCAA CAGAACTCTA GTCCCAAGG GGTGGCCCTA 240
GCCAAGCCA GCTACCGTCC TGTCTCTGTC TTCTGTGCCG GGGCTTGCT CTCATTTCTT 300
YCTCATTTCC TCAATTAAAG AGTGTATTAT GAGCACTCTT TGTGTGCGAG CATTGGCTCA 360
GGTCTTAGC AATCMTGCT GAGCTGTGTA TCCAGGCTAA AGCTAATTAA CTCACAGAAA 420
ATCAGTAACA ACATAATTAC AGTGTGTTG TGCCAGTCA TCACTGTAAAT CCCAGCACTT 480
TTGGGAGCCA AGGTGGGAG ATCAATTGAG GCCAGAGTTT GAAMCAGCTT AGGTAACTTA 540
GTGAGACCCC CTATCTCTAC AAAAATTTT AAACATTAGC TGGGCATGCT GGTATGTCTT 600
NACAGCTCTA GCTACTCAG AGGCTGAGGC AGCAGATCA CTTGATGTTCA AGAGTTTCAAG 660
GTAGCACTAA GCTACAACTA CACCACTCA TGCCAGACTG GGTGACAGAG GAGAACTTCA 720
TCTCTTTAAA ACATAATTAT AATAATTACA GACTCAGAAA ATCCAGTCAA AGAATAATAC 780
AGGTTGGCCA GGTGAGGTGG CTCATGCTGT TAATCCAGC ACTTTGGGAG GCCAAGATGG 840
GAAGATTGCT TTGAGACACC AAGTTTGAGA CCAGC 875

(2) INFORMATION FOR SEQ ID NO: 43:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 843 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

50 CCCAGCGGCT CCGATTCGTC CTTCCCTCAC TTCCAGAGGCT GGGCAGAGCT GAATACCCAG 60
60 AGAGGACAAA GTAAAGGCTCC AGTTCACAAA CATTCTAAGG ATGATATCAT CCACGTGTCT 120

198

5 CACCTGACAG TTACAGAGA AACCCGACC CAGATGCAC GTGCTGTCTT ATGGCAACAC 180
TCAGCCGAGA GTGCTCAGGT CCGGCCACAC TCGGGCTGTG CTTGGTCTG CCAATGGAAAT 240
CCTCAGGACT TTCTCAGCTT CCTTAATGCG AGAAGCCCTT TTACAGCAG AGATTATCCG 300
TTTCTCTGAA AATAGCCGAA CTGAGCTTTT CTTGAGCTTA TATGAGAGT CTCTAGACAG 360
TGGCAGCCT CAGAAAGCCC AGAGCTTGT GATAGCTCCC ACCCTGCTCG GCTCAGATCT 420
TCCCATTTT TTCTCTCTCG CACTAACCTC ACCTTTTGT TTTTGTGTGT TGTGTGTGT 480
TTTGTTTTGG CAGAGTTGGA TTACAGAAC TCTTATGAAA TTGAATATAT GAGAGAAAT 540
GGCTCTCTCT TACTGTAAAG TTGCTCTGCC TCGGCCACT TAGGGGACTC GCTTTCTGTC 600
CTTCAGGGCC CTCCTCCCTT GTCCAGAGTG TCTCTGGGAG CTCAGACCCC AAATCGAGTG 660
TTTTCTGTGT ACACAGCTTC CCGGTTGCAC AGCAATCATG GACTGGGGCT GGGGGGTTGA 720
GGTTTGTACT CATTCACTTT CATTGACAT TTTCAGCGAG AAATGCTAG ATATACATTA 780
GAGCTCTGTC AGAATTACTT TCTTAGAGTG AGAAGAGCT AGAGATTCTT TTAACAAAAA 840
AAA 843

30 (2) INFORMATION FOR SEQ ID NO: 44:

35 (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 489 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

40 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

45 CTCTTAGGCT TTGAGCATT TTGTCTGTG CTCCTGATC TTCAGTGCAC CACCATGAAG 60
TTCTTAGCAG TCTGTACT CTGGGAGTT TCAATCTTC TGGTCTCTGC CCAGATTCG 120
ACAAGAGTGT CTCAGCTGA CAGTATCCA GCTACTGCTC CTGCTGATGA TGAGCCCTT 180
GATGCTGAAA CCACTGCTGC TGCAACCACT GCGACCACTG CTGCTCTTAC CACTGCNACC 240
ACCGCTGCTT CTACCACTGC TCGTAAGAC ATTCCAGTTT TACCCAAATG GGTTCGGGAT 300
CTCCGAATG GTAGAGTGTG TCCCTGAGAT GGAATCAGCT TGAGTCTTCT GCAATGTGTC 360
ACACTATTC ATGCTCTCTG TGAATTCAT CAACTACTTA CTTTGGCTTAC GATATCCCTT 420
55 TTATCTTAA TCAGTTTATT TTCTTTCAA TAAAAATAA CTATGAGCAA CAAAAA 480
AAAAAAA 489

60

(2) INFORMATION FOR SEQ ID NO: 45:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 514 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

60 GAGAGAGTGT GTATCTATGA TATATCTGT GTTCACTAT ATATTTTGA CATATGAA
120 CAGCTTCCA TCTATCAG GAACTCACT GGTCTGGGT CTCCCAAGT CCGAGTGGT
180 GCGTTTGACA GGTAGAGAGA TCGAGTCTG CAGGCTATTT TGTTTTGTG TACAAAACTG
240 TCTTTTCCCT TTTTCCCTCC AGCTATATCA GCAATATCCC TGTAGAGTGG TTTCTACAAAT
300 CTCTGGAGAC TGGGCTGAGG CAGGGGCTTC GCTCTATCT CCGTAAACAT ACTGTCTTCC
360 TTTTCCCTTG CCACTTACGA GTTATCCCC CAGCTATGCC TTTCTCCCTG CTCCCTTGGC
420 CTGGCATATA TTTTGCTCTA TTTATCTGAC AATATTAACA TTAAACTATC AATGTAGAAA
480 AAAAAAACA AAAAACTCAA GGGGGGGGCG GTACCCAAAT CCCCCATAT TGAATCTATAT
514 TACATATCAC TGGGCTCTGCT TTTTACAGG TGTGTATAGG GAAAACTGG GCGT

(2) INFORMATION FOR SEQ ID NO: 46:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1174 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

60 GCGACGAGTC CCGAGTGAAC TCAACCCGCG CCGACCACTG GCGGTGGTGG CTGTGTCTCA
120 GCTTGTGTGT TGAATGCAAT GTTCTTACGA TCCCTCTCCC GTTCTTCTCA TCTCTTCAAT
180 CCAAGGTGCT GCGAAGGAC GCGAGCAAGG AATCAAGAT GAGAGCGAG ATTCAGAGCA
240 TGAAGCAGGA GCTTCTCACA GTCAACATGA TGAAGAGATT TCCAGATAT GCGAGGCTGG
300 AAGAAAGAT CACAGAGATG AGCGATAGAC TCAAAAAACA TGTGAAGCT CCGACAGACT
360 AATTTAGCCA GATTAAGTGG GTGATTAATG TCGCTTTCTA CATTATCAG GCTTCCCTGA
420 TGAATCACT CATTGGGAG TATTAATCTG TCCCTTGGC TGTCTGTCCG AGTAAATGGA
480 TTAACCTCTT AAGCCGCTG GTAGCTTTC CTACATAGAT AGCAGTGGT GTTGGAATTA
540 CCGTTGCAAT TTTATGCTGT AAGAAAGTGG TCGCTATGTG GCTTCAATCG TTCACTGAAA

600 CAGAGAGTGT GATPACGCC GAGGCTTAAA AAGCGATTT CCGTCTCTCA GCTTAAATTC

660 TGAATTAAC TGTTTTGTTT TTTTAAAC AAGATGCT ATTTATGAT TTTTTTTTG

720 TTTGAATGAT TTTGTCTGG ACTTATCAG AGAGCTTAT AAGATCAGG ATTTCTACA

780 CCGTCAATGG AACCAAGAAA GTCCAGTTTA TGAAGATAT GTACATGCA AACCCGCTCT

840 CGATCTGAC GAATATGAA ATGTTTAGG AGATTCAT GCTGACAT GTGATTTGCC

900 CTTTATATGA TTTTGTGCT ATTCGCACT GGAATGTCA AATTTCTA CAATTTAG

960 TAAATCTTT GAACTTAG TCGTGTTTT AATCCAGTT AGAAATGAC TTAATTTTAA

1020 TACACTTACT AAAAAATGCA AATTTCTTC TTTAATGCA TTCAATTAGG TTAAAGAAC

1080 AACACTAAT GACATGGGT GGGCTTTTC TCCCTTTGT TAAATGTCA TTTGTTGAC

1140 AAGATGTGA TATATTAAT TACTACTG AGGCTTTTAA TTTTCAATTA CAGTGTTCG

1200 TAAATGTAT CAGAGAGCA TATGCACTG ATTTATGCT AACTGTGTT TGTATTTAA

1260 AGCATTTTGA ATGAAATGTA TTTTATAGG ATTTAATAT TATGCTTTT AGAATGAC

1320 ACGAAACAA AACCTTAAA GTCCAGATTA ATCTAGCA ATPACCTG TGCCATACA

1374 AGTATATTC TATTAATAT TCCTTAAAC AAAAAAAAAA AAAA

(2) INFORMATION FOR SEQ ID NO: 47:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 596 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

60 GAATGAGCA CGAATTAAT TGAATGAA AGAATCAGG TTCAAGTTTA TTTCAATTAAT
120 AAGTATGTA AATCATGTC GTTCAATAG CCGTCAATTT GTAACTTGA AATGTGAAT
180 AATTAAGCTA GTCAATATA TTTCAACTG CTACCAATG GACCAAGG ATTTATATTT
240 ACAATACAA TCAATCTGCG TGAAGATAT ATTAAGACA ATTTCAATGA AGAAATGTC
300 TTCAATTTT CCGTTTCTCA AATCTGCAAT GGTAACTA ATTAATTAGG TCAATGAAA
360 ACGCTTATTA TTAATTAGG TTCAATGTA GAACTGTCG AGAAAAATTA TCGTTTATTA
420 TATTTCTTGA AATTAACATA AATTCATTA ATTTCAAT CATGAAAT TACCTTAAAA
480 TTGAAAAA TGTGCATTC TACTATATA ACGATATA ATTCATAT CAATCTTTT
540 TTTTTTTTT TGTTTGAT TGAATCTG CTCTGTGCG CAGGCTGGC AACAGAGAG
596 GACCTGTCT TAAATTAATA AAAAAAAAAA AATCTAGG GGGCCCGGT ACCCTA

201

202

5 (2) INFORMATION FOR SEQ ID NO: 48:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 851 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

15 CACATGAGA CACACAGTGG TGAAGGCC TTCCGCTGG CCCCTGTCC TTATGCTCT
CCTCATCTGG ATACCTGTA AGGCCACGAG CGCTTCATA CAGGAGAGA GCCCTACAG
TGGCCCTCTT GCCCTTATGC CTGTGCAAT CTGCCAACC TCAGGCTCA TGGTGCATC
20 CACTCTGGTG ACAAACCTTT TCGGTGAGC CTTTGCACCT ACAAAGTCA CCAAGGATG
AACTCTAAC CTCACATGCT GGGCAGCA GGGCAGAGC CTTTGGCTGT GCCACCTGG
CCTATATACC GGGCCACTGG GACAACCTACA AGGCCACCA GAAGTGCAT GCCACGCTG
25 GGGCAGGAG GCTGTGCTC TCTGCTCTG AGGCTTGGG CCCACTCAT AGCCACCTT
CTGTTTGGG CTCCTGGGC CCACAGCC TGGGACTGC TGGCAGCGG GCTGTCCACA
CAGACTCAT CTGACTAGG TCTCTCTCC CCATGTTTTA TACAGAGGA CCAGAGCCA
30 CCTTTTCTC CCCCCTGGC CAGGGCTCC ACACAGACTA AGTTAGGCAC TATATAGAAC
AGCCACACC CATGGGGGG GGGGCCATA TGGACAGGG GACCTTGGCT TGACTAGGC
35 ACTTCAGAG CTCAGTAGA AGGGCCCTGT ATTCACTCC ACTGCCACA GGGGCTGTG
ACAAACGGC TGGGGGACTG CCAAGCTCC CACTGTGTTA TTTAACTTAT TTCAGTCTT
40 TATATTAAG GAACACTAA CAAGGCATG TCTATGCTCA ATTGGCATG GCAAGCATT
TGGCCTTACC C

(2) INFORMATION FOR SEQ ID NO: 49:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

50 GTGAATGAA AACAGCTTTT TTATAGCTTT TAGCTTCTGA GTTTGGAGT TTGGGGGCTC
60 TTATGTTTGT TTTCCTCTT CTGTTCTTG GAGGAGATT GAGGCTTTTC TTAGTGCAT

ACACAGACC AGGTGAACAC GCTGACTGTG AACCTGCGCT GTATCCGGAG CTGTGCTGG 180
CACTGAGGGG ATCCACAAA ATTAGGAGAG GATCTCTGCT CCCAGCTCT ACTTCTCTTA 240
CCTCAACAGG GGTCCAGGCT GCAGTGAAT CAGTCTTGG CCTTGGGTG AGGATTCATG 300
GATGATGAA AGCTAGACT GATGGGAGG CATTATGACT AAATAGGCC AGCCTCTTTC 360
CCTTCAGCT CTGTCTTAGG AGCATAGGG GGAATCTGA GTAGAGCTG ACTCAGTTTT 420
TTGCTTATGA TTTGTAAAG CCGTCTTGG GTCAATAGA AANTAGGGT GATGAGGGG 480
GAGAGCCCA GGACTGGAG AATGCACT GCCCCAGGG TTTTACCNA GGAATTTCAA 540
GAAAACTGG AGTAAGAAT AAAGCCCCAG AGGATTTAAT TATCTGTTT TGCANAAG 600
CCTCCATCG CAGTACGCC CAGCCTTGA GGGCGGAGT CTCATGGCC CTGTGCTCTG 660
CTGTCTCTTC AGCCATGCC CACGAGTAC CTCTGACT GAGAGCGGC TCAAAAGCTG 720
ATTAGAAAG GAGGAGCAC TTGTGACTTT GTTGACTCT GTGACTCACT TCCTGCTCA 780
CACTTGTTT GACTACTGG ACTTTCACCT GCTTTCTCT AGCTCAGCCA AGCAGACAG 840
TCCCACTGA AGAGTCTGT ACAGTGACA CCGGGCGGG CAGCAGGAC ACAGATGCAG 900
CCACGTAAG GCTCCATCAG CACTGGGTCA GTATGGCAA CAGATGGCC AAGATGGCT 960
CTAGAACAT CTGTCCATGC GTCACTCCC CAGTTTTT TTTTAGCTTT GCTTCCAGG 1020
AGTGACAGC ATCACAAATA CCNCATCT CTCTACTCT CCACATACC AGATSTACA 1080
CTGTTGTTT TTCATGAG GTGAATGTT CAGAGATGG GGGCATCTG GTTGTTAAG 1140
AACTTACACT GGGGAGCTTT ACTTTCGCT GTCAAAATG TGACTACATG TTCTCCAGT 1200
TAGCCACCA TGCAAACATC AGTGTCTTC TAGCTTANC CGAAGAGAA ACCAGTCCA 1260
GGGATGAT GGTGGTCTCC CCACTCCGG CAGCACTTTA GGCAGCCAT AAGCTATGG 1320
AGAATGTGA CGCTCACCTT GCTCCGTGAC GTTCTGACC TACACATAA ACAGGAGAA 1380
GCCAGTGAC GBAACGCTC TAGGATGAC AAGTCAAGT AAGATGTGC TTTATATAC 1440
CAGAAATAT GGGCTTGGC TAAGTGGTG TCTCTTACC TGGCGGGTC ATTCCCAAC 1500
AAACACCCA TACTAAGGAG CCAAGGCCA CCTGACAT CACCTTTTCT TTGACCATCT 1560
GAGTCTTGG CCAACTTAAG GAGGONCA CACAGTGGTG CAGGCAAT TTCCAGGTA 1620
GCTGTCCCTG CCTTTTGTG CCAAGCTAG TGTATGCTC AACACNGEC CAGGCTCT 1680
GGGGCACTGA CTTGAAAGT GGCANAATG AGTTTTACA GCTGTGGCG GAGCAGAGC 1740
CCTTCTTCA TCTAACATC TGAATTTCT TTAAAAAG AAGAAAGGA AAGATTTCA 1800
TAAGCAGTG TGAATGACA GTTTAAGTAC TTACCATTT CTCTTCTTC TTATGAGT 1860
GACTGTGCT GTGATTAAT CATTGTAT TCTGTATGT TCTATGAC TAACATTA 1920

TTAAGTGGT	GTGTATATGT	GTAACTAATG	TAACTGGCTT	TTAAAAATTC	ATTACAAATA	1980
AAATGACTTT	GCTCTGAATA	AAAAAAAATA	AAAAACTGTA			2020

(2) INFORMATION FOR SEQ ID NO: 50:

10

(A) LENGTH: 2432 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15 (b) 10/20/2001: 2:11:00

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

[illegible]

TCATCGAAGC CAGCCAGCCC AAGGCAATGC TGTCTGTGATG TGCCTTTCCC TGCATGCTTT	1332
CCATGTTGGA GCAGAGGTGT GAAAGAAATT TACGTGTGTG TGAATGCCAA ATCAGAGAAC	1380

5 AGAATTTCAT AGCCACAGCT GCCCTGTGT TTAGCTAGA AGGCTTCT ACTTCAGTT 1460

TCATCCACA AAGATTAAA AACTGGTAC ACCACAGGCT TTCGACACAT CCATTGCTTG 1500

GGTTTGCAT TTGACCCAC CCTTCCCTA CCTGAGAGC TTCTTTTGA AACGAGATG 1566

1620

CTGAGCTGG AAGGCATTT GCATGCTCT CTGTTGGGC CTGGGCTGC AGACACACC 168

15 ТООСНННАС ТООСНННАТ ТАБОННОС ТАБОНАНОТ СӨТНННОТ ТУОАНАС 174

ГРМСССНТАС АТГГГГГГГГ ГСГГГГГГГГ ГАГГГГГГГГ ГСГГГГГГГГ 180

TCAGGCGCTC AGTGAAGTT GECTAAGGT TGTGTAAAA ATCAAGAGAA GCTTGAAGA 186

CAATCATGAT GCCATGATTT AGCTGTGCAA CTGACCAAGT CCAAGTTTGA TCAAAACCAA 192

AGCAGACTTTC CTCATGTGCT CTCAGCATGT GCGATGTTT CTCAGCTTCC TAGACCTTCC 198

25 ТТАССТСАТ ГТТТСТСАТ ТАСАТТТТ ГААТССАС ТТТАСТСТ СААСТСТАА 204

GGAAGCTTTC TCTTACACC TTGGCTTGG AATATGCCA GAGAGAAAT TTGGCTTTT 210

TTTTCTTAT GAGACAGCA CAGTTGCTT TCICATGTC CAAGTCGAG AGCACCAGAC 218

30

CCGACATC TCCTCTGGA AGATTACT GTCATGAGC AGCAGAGCT GAGTGTCTC 22

СРЕДСТВА СРЕДСТВ ТРАНСФОРМАЦИИ ТРАНСФОРМАЦИИ ТРАНСФОРМАЦИИ

27 TGGCCCTGGGA TTAAATCAGT TACAGGGCCAG AGTCTCTTG GAGGAGCTTC AACCTCTTAA

CCTCCTATNGA ACCCTGTGTA G CCAAAATAGA ATTCTTAAAA TACCCUAGG AGCCCCAAA

34

AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA

40

(2) INFORMATION FOR SEQ ID NO: 51:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2340 base pairs

(C) STRANDEDNESS: doubt

(D) TOPOLOGY: linear

Figure 1

(x1) SEQUENCE DESCRIPTION: SEQ. ID NO: 51:

60	GAACCTGATTC	CTTCAGACGCT	TCCCTCCGATT	GCTACACCCG	AGCTCTAAACG	AAAGATGAGAA	240
55	ATTATGTTATGC	GGACAAACCG	CGCGGCTCTGG	CGGAGNATAGC	TCCCTCTCAGC	CGCGGAGAACG	120
60	GAACCTGATTC	CTTCAGACGCT	TCCCTCCGATT	GCTACACCCG	AGCTCTAAACG	AAAGATGAGAA	240

205

5
6000CAGCTT GTTATTAA GACATCTCA AATGTACAT GTTGTGTGTT GAGTGTGGA
TCTTTTAT CTTCAAGTAT ATTTATTA CTGAGATG TGACATGAA AATATGCTT
ATGTGGACC TGAACATGTA AGAGAGCTC AGAATATGTC TCAGCAGTC TTGAGAGG
AATGTGTC CAAGTTGCC AAGATATCAA TGGCCCTGTT ATTGTAGAC AGGTATAGCG
TGGACTTACT CCTTTTGT CAGAGGCC CCAGAGACAG TGAGCTGAG TCCAGTAGC
ATCTCTCTTT TGGTTCGCG AATTTCTCCA GTAAAGTCCA GACCTCTGTC GAACCTTGC
CAGAGCAGGA CCTCCCTGAA CACTTGAAG CCAGAGCTG TCGCCGCTGT GTGTATTG
GAGCGGAGG AATAGTGCAC GATTTAGAC TGGGCCACAC CTTGACACAG TTGATGTTG
TGTAAAGTT AAACAGTGCA CAGTTGAGG GATATTCAGA ACATGTTGGA AATATACTA
20 CTATAAGGAT GACTTATCCA GAGGCCAC CACTGTCTGA CCTTGATAT TATTCGAATC
ACTTATTGT TCGTGTTTTA TTAAAGATG TTGTTTCAA CTGCTTCAA CCATGTGTA
AAGAGAAC CTGCGATTC TGGTACGAC TCTTCTTTTG GAAGCAGTG CCAGAAAAA
TCCCACTGCA GCGAATAT TTAGGATTT TGAATCCAGT TATCATCAA GAGACTGCTT
1020 TTGACATCC TTCAATGACT AGAGCTCAG TCAAGTTTCT GGGGCGCAG ATGAGAGCT
30 CCCCACATC GTGTCTATG CCGTGTCTT AGCCACAT CTGTGATG AGTCACTTT
GCGCGTTT GGTATGACC TCATCAACC CAGACACCT TTCACTACT TCGACATGA
ATGATGCT GCTATGACT TTGAGACAT GGTATGTC ACAGCGAA CCAAGTTCTT
35 CTTAAGCTG GTCAAGAGG GAGTGTGAA AGATCTAGT GAGCGATG ATCTGATTT
TTGACACAG AAGACTCAG TTGAATGCG AACTCTAAT CTGAGACTG TTTTTCAGC
40 CTTCTTGAT GTATTCTCC ATCTGCGAG TACTTTGAG TCGAGCTCAT GTTTTAACT
TTTAATTTAA AAGACAAAA AATTTTATG CTCTTCCAC TTTTTTTTC CTATTATTT
GAGTCAAGT TTTGTTTTG CACACATTT TGTAAATGA ACTTAAGAT TGAATGGA
45 AGACTCTCA AAGAAATG TATGACGA TGTGTGTTG ATTTTAAAG AAGTAAATTA
ATTTGTAAT CTTCTGCTG TTACACTGC ACATGATTA CAGTAACTA ATGAGAGGA
50 GAGGGAGGT CACTCTTTTG ATGTTGGCC TGAACCTCAT TCTGTGCC TCTGCTGTC
CTTGTGTGTA CCGAGGAGG ATCACTCC AGATGAGCT GCTCGTAGC TCTGCTGTC
ATACTGGTC TCGATGAG CCGCTGAG CCGTGGCTG TTGAGAGG TGAACCTT
55 TCTGTTG TCTGCTTCT CTTGAAGAC TCGAGACCA ACCAGGAG CTGTCTGGA
GTCCTGCT CCGAGGGA CATGATCT GTGACCTG ACACTGTGA AGCAGCTG
60 GCTACAGAA ACCACATCT TCCAGCAT TATTAATTT CTTGAATTC TTGGGGAATTT

206

2100 TTTACTGCC TTTCAGCA CTTAGTGT AGATCAAG TGTTCAGTG TCTGTCTGAG
2160 GTGACTTAAA AATACAGAC AATACTTCTA TTATCCAGG TCATGGAGA GTACACCTT
5 2220 TCCAGGATA ATGTTTGG AAACACTGA ATGAATCTT CCAAGTATTA TAAATGTGT
2280 AATTAAAAA AAGAACTTT TTGAATGCC TACTGCGGT GTATCCAGG CAGTGTGCCA
2340 GTTTAAAG ATGAAGA AATAAACTT TTGAGGAAA AAAAAAAA AAAAACTGA
15 (2) INFORMATION FOR SEQ ID NO: 52:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 601 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 52:
25 AATAGGGAG ACTGAGCTG ACCGTAGCC AGGCGGGG AGAGGACG CCGGACAGA
120 CTGAGCGGC GCGGAGAAC CACTCAGG TTCCCGCG CTTTCCCTT GAAANCTAGG
180 CTTTCCCTT TCCGTTGG CCGGAGAG AATCTGAC TCTGCGACT TACGCGGAC
240 TAAAGATT TCAAGTAG GAGAACGA TCAAGCAT CCGGAGAG GAGGACCA
300 GAGCGGCT CCACTGCG TCCCTGCG CCACTAAT CCGGCGCA ACCGAGCTT
360 TCTTAACA CCTACTTC TCTCTCTT TTAGCATG TGGCTGATG GACATCTGA
420 CAGACAGG ACTGACAT CCAATCTGC CCGCCCGCA CCGGAGAC TACAGTGTTC
480 TCGATTGAC CATGACCTG GATGTGAAA CTGTATGCT TTTTGGCTG ATTGTATGC
540 TCTGCTTGT CATGTGATA CTATGTTT TCTGGAAC GCGTGAATG GAGTCCAGC
600 ACCTGAGTG TCGGAACTC TCGTTTGT TTATCCGA GAGCGACCA GAGAAAAA
601 A

50 (2) INFORMATION FOR SEQ ID NO: 53:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

60 CTCTGCGGA ATTGCGACG AGAGATGTA CTTTAAGAG GTAATTAGT TCTTAAGTG

207

GAATTAACATC TTTCCTCTTA CACTGAGACT GGGTTCCTCT GGGATGGTTC AATTCCCAAG 120
AAGATGAGTT GTTATTAACAC AATGCTGCTT CTCTATATTT GCGCTTTTTC TTTCACACAA 180
CTGGGTCCCC TTCTGTCTTC CTACAGTTTC TTGATGCTGC ATGAGGCACT CATGAGAAC 240
CACTGATATC AGCTGCTTGA TCTGATATTT CCGAGCAAC AAGACCAAGT GCTAAATATA 300
ACTCTTTTA ATAAATTAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 359

15 (2) INFORMATION FOR SEQ ID NO: 54:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1141 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

GGCAAGACT GGTAGAGGCT GAGATATGCG TCCCGCGGCG GGGTTCGGA GCAATGCGGA 60
CCCCGAGAG TGTATTAAGA CGAGACAGA GCGCGAAAT ACGTTCCAA CTGACAGATG 120
ATTGATATCC AAGACAGAT GCTTGGGCA GCAATTGAGC TTCTTATCT GCGAGAAAT 180
AAGCCTGTT ACCTGCTGGA TATGAGCTT GCACTGCGGC TGAATGAGG TTATCTTCA 240
GATGAGAGGC ACTATGAGGT GGGCTGTGAT ATCAGCCCTG CCAATCTGGA TGAAGCTGTG 300
GATCGAGAGA TGAAGAGAGA CTTGCTGCTG GGGAGATGAG GCAAGGCAAT CCGATTCAGG 360
CGAGGCAAT TTGATGAGTT CATACAGATT TCTGCTGTC AGTGGCTTTC TAAATCTTAC 420
AAGATGCTG AAAACCTGCG CAAGGCGCTG TACTGCTTT TTGCTTCTCT TTTTTCGTCT 480
CTCTCCGCGG GATCCGAGCG TTCTCTCAG CTGTACCTTG AATATCTAGA GCAATTGAGG 540
CTGATCTCAA CCGAGGCAAC AAAAGCAAGG TTCTCGGCTG GCAATGATAT AAGATACCTT 600
AAGATGCTGA AAGCAAAATA ATTCTACCTC TGCTATGTTT CTGGGCTTTC GACCTTTTAA 660
CGAGAGGGCG TGAATGAAA TCGAGATGAA GTTGAACCA GGAATGCTCT GTTTCACAAAT 720
GAGAGGTTCG CATTGAAGAT GTTCAGGCGG GGAATGCTGA GGAATGCTG GCAATGAGTTG 780
CTGAGAGAAA AAGAGCGGCA GAGGCGCAG GCGAGGAGG TCGAGCTTGA CACCACTTAC 840
ACCGGCGGCA ACGGCAAGCC CCGCTTTTAA GTACACAGCG GGTTCGAGAA AAGGACTTGC 900
CTCTGCACTT TTCTATATTC TTGAGCTTAC AAGATGATAT TTGAGAAAG TTCTAAAGTT 960
ATTAATAATG TTTCGAGCT AAAAATAAG TTCTCTGCGC GGGGCGTGGT GACTCTGACCC 1020
TGTAAATCCA GACACTTGGG AAGCTGAGGT GGAAGATCTA TTGAGGCGCA GGAATTTGAG 1080

208

ACCTGCTGCG GAGATATAT GAAACTTCTT TTCCAGGAGG AAAAAAAAA AAAAAAAAA 1140
A 1141

5 (2) INFORMATION FOR SEQ ID NO: 55:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1560 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15 (14) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

TTCTTCTCTG GGGCGTTCG GTTGCAGAG GATGCGGAAA GCGGACTCT GGGCTGCAAG 60
TACTAGAAAT TTAGTGGCTT CACCGAGAG TTGGCAGAG GATGGCTTTC GAGAGCTTAT 120
AGCCGCGAGA TTAAAGGCT GTGGTTTCCA GAGAGCAAC AACTATCATTA TTTCGCAAC 180
CACTTAATCT GACCTTCGAT TCCACAGAT ATGATTAATC TGGAGAAAAC AAGTTCCAG 240
AGCTGAAAA GTTTTCCAG AAGCTGATG GTTGCGGCTT CTACTGAGAA CGAGGCGTTC 300
CTAGCAAAAT GCTTACCGCG AACACATGCG GGTGAGCTGT GCGAGGAAAC ATCTACTGCC 360
TGATGCGGCT GAACTAGCT TCCAGGCCA AAAACAAATG AATTAGGCTG CAGAGACTG 420
GTGTGTGTTT TCCGATTAAC CATTGAAAT TCTTTTTCGA TTGTTAATAT AAAATTTTTT 480
TTTTTACTTG GATGGCTTAA CATTTTTCCA AAAAAAAAA GAAATATGTA AAGATGATGT 540
TTGGTTGTTT TATGAAATGC AATGAGCTTG TCGAGCTTCA TTCCAGAGTT AAAGCAATG 600
TTTAAAGAAA CGTTCCTTTC CTCTGTGTTT GTGCTCTGTA TTTCCTTGA GATTCTGAAAT 660
GAAAGCTGAA CACAGCTTG TTAAATGAGG TGTGTGCTGA GGAATCTGAG GACTTGAGAT 720
TGCATTTTTC AACATGGGCT GCGAGAGCTT TTCTGATTT GCAATGTGGCT ATCGAAAGAA 780
CAGAGAACCC AAGGTCAATG GCAATGAATG AAGATTTTGA GTTGAATCAC TCGCGAAATTT 840
TTTCAATTTT GCAATTAAT GTTAAATTTA TGTAGCTTTC CTCTATTTTCT TGGCGAGTTA 900
AATTCAAAAG GTTATTTGCG TCAATCTCTA TCTTTAGTGA AATCTATAT GTTATTTGTT 960
GTAATTAATC CAGCTGAGGA ACGAGAAATA CCTGTTTAT GTTCACTTT AAGCTGCTGT 1020
CTGTGTTGTT AATGAGCTG TCCAGCAAAAT TCTCTTTTAT CTTTAAABAA TGTATAGACT 1080
TTAAATTTTG AATTATTTTG ACTGTGAAT AATATCATGA ATGAAAAAT TTAAATTTGA 1140
AGTCTTTTGA ATGAGCTTTC AAGATTAAT CAGAGACCA GCAAGATTT AAATGAGAT 1200
CTAATTTCT TCTGTGTAAT TACGACAGAG ATTAATCTTA TAAATATGCTC TTAAAGCTTA 1260
GTAAATAGAG GTTAAATGCA ATGATGATAT ACCAGAGAT TTTTAAATAA AACTGCTGCG 1320

209

CCCCWAGTG AAGGTACTT GTAACACA GTTCATTAG AACTTAATTT CTTTTCCTGT 1380
CATGTTGGK AGACTTCACT TACCCTATAT TAATTTTGA AAAAGTGGG ATTTTATAT 1440
ATATGAAGG ATAGTTTGTG TTTTACCATA GCACAGAACA GTGACTCTTT GCTCAGGATA 1500
AGATGTGGT ATTTGAAAT ACTATAGTA GCTTGCAT GTACTCTCTC TCTCTCTC 1560

10

(2) INFORMATION FOR SEQ ID NO: 56:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1507 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

GGAAGCGAGA GCGAGCGTG GAGAGCGAG CGAAGCTGGA TAACAGGGG CCGATGATGT 60
GCGAGACATC AGTTCTCTG CTTCTGTTC TACTGAGGA CCGGGCCGAG CGGAGCCAT 120
CCGAGAGCC AGGCCCTCAT GCGCAGGGG GGTGTGACCA GCGGGCCCC CTGAGCGAG 180
CTCCCATGTA TGAGCCGAC GGAAGCTTTC AGTACGACCA TGAGGCTTTC CTGGAGCGG 240
AAGTGGCCAA GGAATTCGAC CAATCAACC CAGAGGAAAG CAGAGCCCGT CTGGGCGGA 300
TGTGTGACCG CATGAGCCG CCGGGGAGCG GCGAGCGCTG GTGTGCTG GCGGACCTTC 360
GCGGTGAT CCGGCACAG CAGCAGCGCG ACATACGGA CTCGTGAGC GCGGCTTGG 420
ACACGTACGA CAGGACCG GAGCGCGTG TGGTTTGGG GAGCTTGGC AACCCACTT 480
ATGGCCACTA CCGGCCCGGT GAGGATTTTC ATGACCTGGA GATGCGAGG ACCTACAAA 540
AGATCTGCG TCGGAGCGAG CCGGTTTTC CGGTGCGGA CCAAGATGG GACTCGATG 600
CCACTCGAGA GAGCTGACA GCTTCTCTG ACCCGAGGA GTTCCCTCAC ATGCGGACA 660
TCTGTATTC TGAACCTG GAGGACTGG ACAGAACGA AGATGCTAT GTCCAGGTGG 720
AGAGTACAT CCGGATCTG TACTCAGCG AGCTTGGGA GAGGAGCG CCGTGGGTG 780
AGACCGAGG CAGCAGTTC CCGGACTTC GGGATCTGA CAGGATGGG CACCTGATG 840
GAGGTGAGT GGGCAGCTG GTGCTGCCC CTGCCCCAGA CAGGCCCTG GTGAGAGCA 900
ACCACCTGCT GCACGARAG GACAGGACA AGGATGGCG GCTGACAAA GCGGAATCC 960
TGGGTATTC GACATGTTT GTGGGCGTC AGGCCACGA CTATGGYAG GACTGAGCC 1020
GCGACAGGA TGAGCTGTA GCGCGAGCA CCGCCAGAG CCGTCCAGG CCGCAGATG 1080
ACCGAGAG GGGGCCCTGT GGTCTGGCC CTTCTCTGTC CAGGCCCGC AGGAGCGGA 1140

60

210

TGAGTCCA GCAATCTCC TCCCTCGG CTCTCAGGA CCCCCGGT CCGCTTCTGT 1200
CCCTGTACA CCCCAGCC CAGGAGGGG CTGTCTAGT CCGAGAGT AGCAGTACC 1360
TATTTCTGAC TGATCTCCC AGCCCCAGC CAGGACCTT NCGCCCCAG CTCAGCTCTA 1320
AGACCGCCC CAGCCCTCC AGCTCCAAAT CTGAGCTCC ACCCATAGA CTGAACCTCC 1380
CTGCCCCA GCGCTCTCT GCGTGGCTG GCGTGGACA CCGCTCTCT GCGAGAGGC 1440
AATTAAGCC AGCGCCGGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1500
AAAAAAN 1507

15

(2) INFORMATION FOR SEQ ID NO: 57:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 450 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

GAATTCGGA CAGCAGTGT CCAACTGT AGCTGTGCC TCCAGGTTT CCAAGTGGT 60
GCGTACCGG GTCTGAGAG AGATGTCTG GCTCGGGA TGGGSCAGA TCTTCTGCG 120
AGTTTCTTC TCCCTCTTC TCACTCAAT GTTATCAGC TTCTCAGGA ATGTTTAT 180
CCAGAGCCC AGAACATC AGAACCCAG AGATGGAT RAGAGGAT GTCTGTAAA 240
GAGAGTGT CATTGTCCA CAGAGATGA GAATATATG ATGATAGAT AATTGAAAG 300
AGATCTCCA GAAAGAGCAG AAGGAGTTT CTTCATGCG TTCTTCAGG ATTTTATCA 360
TCTTACAGC CTCTTTAGA ATGATGAC TTCCAAATC CCGAGATTA AATTTTAAA 420
TTCTATAAA CATTTTTTCG AGTAAAAAA 450

45

(2) INFORMATION FOR SEQ ID NO: 58:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1147 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

GCCACGAGC CCAATGAGCA GAGGAGGCC AGGTGGGAA GCTCTGGGA AGAGCAGCA 60
GACTGAGAC TGGGCTGCTT GAGTCTGAG TCAATATCA GAATCTCGG GCTCTCTGG 120

60

TCATCTTAT CATTCAGATT GAAGATTGC TTCTTCAG TCATCTGCCT CTTCATCTTA	180
CTCTCTCTGG CTCTCATTTTC AGATCCCAT CTCTCATGATG AAAATCTCAA GAGAGCTTT	240
GTCTGTGACA CCGCTCTCTGC CATCTGCAC TACATTCGCC ACTGACGAAA TCACCCGAAA	300
TTACTGTGCG GAGGCTATTTC CCGTGCATAC TGCATATCA TCCGCTCTCT CCGTACACAC	360
ACCAATCATG TGGCTCTGAA GACACAGAG AACCATCTCA TTCTCATCTAT GTCTCTCCCTG	420
AACAAAGAG AACGCGCTG GTACTGTGTG GGCATTCAG GGGACTTTGC GAGGATGAC	480
ATGTGATTTTA CAGACTGAT TTCTACTGAC GACAAAGGAA CTTGCGCAT GTCTTTGCTC	540
TTGGAAAGAC TATTCAGCGAC AAAACCGAAA CCTGCAGGC TCCAAAGTT GTCCCGAAG	600
CTGACCGCTC CAGACGTCC ATTCTCATCA TTTCATATCT GATCAGCGGT TTGGAAATCA	660
TTCTGTATAT CATCTATTTC AACAAAGCA GAGAAATCA AAGGATTTGA AGGGTAGACA	720
AACATTTGAA GCGCTCTCTG CCGTCTCTGA CTCCAAAGG AATGCGCTCT ACTGACACGA	780
TTTGACTGAA GATTTTTTTA ATTATTTCA TTAAGTGATG GTACAAACGA ATTAATCACA	840
TTGACATGCG CCGCACACT CAGACATGA TTCTGATCTC CCGGAAATTC TGAAGCTCC	900
TTTATCTTTC AACAAATCA TTTCAGGCA GTTAGCAGG GAGATATCA GAGGACTAT	960
GATGACACAC ACCCAAGCA GCGTCCCTTC AAATGATTC TCAAGATCTT AGTCTTTATG	1020
CATTCATCA GTCCAGATG AAGAAAGAT GAGATATTC GATTCGAGAC CAGGAAATCA	1080
CTTGTATTTT GTTAGCCAT AAATCTCTAG CCAATGTGTA ATGAAAAA AAAAAAAA	1140
AAAAAA	1147
(2) INFORMATION FOR SEQ ID NO: 59:	
(1) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 777 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: double	
(D) TOPOLOGY: linear	
(1) SEQUENCE DESCRIPTION: SEQ ID NO: 59:	
GGCAGAGCT CCTCAGAGG GCGTGGGCTC TCCATCTTC CAGATCTCC ACCATGCTT	60
GTTCCTCTAC CCGTAGAGTA GCTCAGCCAT CTTTTACTTG CTTGGCTTAG ATCCATGACA	120
TTTCAATTTCC TCTCTTTTGC ACTGCATCA GTCCCTCATCT GCGCCATCT CTGAGAGAG	180
GAGCATAGAC TTTCAGAGGT GAGCCATTC TCTGGGGTCA CACTATTTTAC ATCCAGACAG	240
GACTCCAGCT CATATGTGCC AATGCAGACA CTCTTATCC ACCTGGGGCC CTGGGCTTTG	300
GACCTGGTTC CTTCAGACAC AAGAGACCG GAGCTGAAA GAGCTTTGCG GTTGTCTCAT	360

AGTCACTGCG CAGAGAGAA CCGTACGCC TCCCAAGCTC CAGAGGAGG GATCAGGCT	420
GCGTGTGAC AACGAGGTAG CAGACATTA ACATCTTGT CACACAGAAA TGAACAGACA	480
ATTTATTTAA AACTTTCGCC TTGGAAATGC TGAATCATTT GAAATTTTAT CCAACGTTT	540
GAAGAGGAA AGAAATGTG AATCTTCCA GCGTGGTCT CCGCTGCGCT GGGCTGCGCC	600
AGCTGTAGG CCGGCTTCTT TTCTGAGCAT TCATCTCATCT GATTTTCACT GAGGCGCAG	660
AAGAGCCCT AACAGGGTAT TACATATCA GCGCTCATCT GCTCTGGGA GAATTTACCA	720
TGAATTCAT GCGTTAAAC AACACAGAG CCGTCTGAG CTTACCTGCG CTTCAGAA	777
(2) INFORMATION FOR SEQ ID NO: 60:	
(1) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 1191 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: double	
(D) TOPOLOGY: linear	
(1) SEQUENCE DESCRIPTION: SEQ ID NO: 60:	
AAGATGATTT TTCTTATCT TCCAAAGCT CAGCATTTTG AAGTTTCTTT TATGAAATG	60
GGGCAAGAA TCAGGGTTGAA AATGATGTA AACAAAGCC ATCTGTGGT CAGCACCCAC	120
TGAGGTGGC CAGCAGATG GCGTCAGATG TTCTCTGAC TGGCCGACGA GCGCAGACA	180
GAGGTCAAT CTAGCGCCCT TGTCTGTGCT GATTCATCA GCGAAATTC GAAACACAA	240
AACTCTTGG AACAGAGAC AATCATGTCT GAGAGAGTA CCGCATCTGCG GCGTATGGC	300
CATCGCATTC GTGAGACAA AAGAGATGAG TTACGAAAT ATCTCCATTA CAAAGACATG	360
CATCTCTGCG AACATGAGAA GTCTCTGGGA CAGACACAA CACTTCAAG CTGCACACAC	420
CTACAGCTT TAAATATCTG AAGAAACAAA ATCAGACAT TAAATCAGCA GAGGGAGAG	480
TAGCTGAGAG CAGCAGAGG CCAATTTAT ATCCACAGAA TTTTATTAAA AATGACTCC	540
CACAGAGGG TGGAGAGAAA GCGACTGAT TTAGAGAGTT CTTCGCTCAG CCAACACATG	600
CAGTATCTTA CAGCTTTTAC AAAGCAGAG AAGTAGAGAG GCGCTGCAT CAGAGACCTC	660
CCCAAGGCC GAGACAGCAG AACAGGTGG TTCTTCTTT TTCCCTCTG GCGTGTGG	720
AATCTTAC AGGTGGCT CTGCTTTG GACATGCT TCATCTCAT CCGCGGTCA	780
AAGATGAT CTGTATCAT TTTCAGGGG AAATTTGAG GACAGGCC GCGTCAATAC	840
GTTCACCCCA CAGGAGATG ATCTGAGAG CTTTAAACA CTAATCTGCG GTGGCTGAGT	900
GCTGTACAA AGCTGCTTTT GTCCAGGGCT GAACACAG CAAAGAGGCA GCGACCTGCC	960

GGAGGCTTGA AGTGGGAGA GATCCCGCA GCGTGCAGG AGCCAGGAG AACCTCCAC 1020
 TGGATCTAA CTGTGGACA GCCAGCGGT GCGCTCTTC ACATGGCTC CAGGCTCCT 1080
 5 CAAAGCCCTT CCGAGCCCT GCAGGAGAG AGCGAGGAG AGCAGAGCA GGGAGGCGAG 1140
 AGGTGGCTG AAAGCTTGG CTGGAATC CTTCAGACA GCTTTAAAGT G 1191

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1580 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 61:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

CCCCGCCCC CGCCACAGAA GGAGTGGCT GTGTCTCCG GCGGAGCCA GAGCGGTTG 60
 GCGCGTTGGA CTGGCAGAG TCGCGGCGG GCGCGGAGG GAGCGAGCC GCGATGCGCT 120
 ACCAGAGCTT CTGTGTGAG CCGATCAGT GCCAGCGCT GAGCAGGAC GCGACCCAGA 180
 TTGCGATCTG CCGCAGAAC CATGAGTGC ATATCTATCA AAGAGCGGT GCCAATGGA 240
 CCAAGTGA CAAGTCAAG GAGCAGAG GCGAGGTGAC AGCATGAC TCGCGCCCG 300
 AGATTAACCG TATTTGACC TCGGCAGAG ACCGCAAGC CTAGGTGAG AGCGTAGG 360
 GCGGAGATG GAAGCCGAG CTGGTCAATC TCGGATCAA CCGGCTGCC CGTGGGTGC 420
 35 GCTGGGCCC CAACGAGAC AGTATTGCT TGGGAGGCG CTCTGTGTG ATCTCATCT
 GTTATTGGA GCGAGAT GACTGTGGG TTTCAGACA CATCAGGAG CCGATCCCT 540
 CCAAGTCTT CAGCTGAC TGGCACCCA ACATGTGCT GTGGCTGCC GCTCTCTGTG 600
 ACTTCAAGTG TCGATCTTT TCAGCTTACA TCAAGGAGT GAGGAGCG CCGGACCCA 660
 45 CCGGCTGGG CTCGAGATG CCGTTTGGG AACTGATTT GCGATTCAG AGTAGTGG
 GCTGGGTACA TGGCTCTGT TTCTAGCCA GCGGAGCGG CTGGCTGAG GTAGCCAG 780
 ACAGACCTT CTGCTGTGCT GATGCGACA AGAAGATGG CTTGCGACT CTGGCTCTG 840
 50 AAACACTACC ACTGCTGGG CTGACTTCA TCAGAGACA CAGCTGTGT GCAAGCGGC
 AGACTCTT CCGGTGCTG TTCACTATG AGCGCCCG GCGATGCTG AGCTTGGG 960
 GCGGCTGGA GTTCTTAAG CAGAGCTGC AGCTGGCTT GAGCGCCCG GAGCGCTTC 1020
 55 AGAAGCTGA CAAGAGCGG AGCTCGAGG GTGGCAAGG TCGCGGCGG GCGTAGACT 1080
 GCGTCAACA GAGAGCGTC AGCAGTCTT CCGTGTCTAG CCGCGCAGG CCGAGTGT 1140
 60 CCGAGTCTG CACCACTGC ATGATGGCG GCATGATAT CTGGATGTG AAGAGCTTG 1200

AGTCAGCTT GAAGGACCT AGATCAAT GACTGTGAG GATATGTG CTTTCATCT 1260
 AGCTCTGG GAGCGGGA GAGGGTCAG GAGGCTAAT GTTGTCTTG CTGAATTTT 1320
 5 CTGGGTACC ATACGAGTT CCGTAGGGG CTGCTCCCTC AAAAGGGAG GGGACAGATG 1380
 GGGAGCTTTT CTTACTATT CAGGAATAC GTGCTTTTTT CTTAATGCT TTCAATTTT 1440
 10 GAAAAAAA AAAATGCC CCAAGCCT ATGCTGTCA TGAAGTCTT CAAAATGTG 1500
 AGTATATA ATGACTGT GTAAAAAAA AAAAAAAA AATGACCT CCGATCTAG 1560
 15 AACTAGCG ACCGTTGGT

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1117 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 62:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

GCGCAGGC CGATCCAG ACAGCTAGA GCTGTGCAA GCGGGGGCC GCGCCCTGG 60
 30 ACCCTCGG CCGGGCGTT TGGCCCTTA GCGCCCGGC GTGCGGCGG TAAAGGCG
 CGAGAGGA GCGCTTGA AATGTCTT CTTCAGGAC CCAAGTTTCT TCACCATGG 180
 35 GATGTGTC ATTGTGAG GAGCCCTGG GCTGTGTC TTGGATATG TGTTCGCAA
 CACAGATG TTCTGTCA AGCCCGAGAA AGCGGCGTG GAGTACCTGG AGGATAGA 300
 CCTCAACA CTGCGAGG ACCCAGGAC TTTCAGGCA AGCAGCTAT GCGAAAAA 360
 40 TCGAGCTGT ATTATGCG TCGGAGGCC AGGCTGTTC CTCTGTGAG AGAAGCTGC
 GATCTGTC TCCCTGAAA GCAATGGA CCAAGTGGC GTCCCTCTT ATCGATGCT 480
 45 AAAGGAGAC ATCAGACTG AGTCAAGA TTTCAGCTT TATTTCAAG GAGAAATCTT
 CCGGATGA AGAAGAT TCTATGTCC ACAGGCGCG AGATGATGT TTATGGATT 600
 50 TATCGCTG GAGTGTGT ACAGTCTT CCGAGCGTG AACGGAGCT TCTCTGAAA
 CCTCGAGA GAAGCTTCA TCTTGGGG AGTTTGTG GTGGATCAG GAAGCAGGG 720
 780 CATCTCTT GCGCCGAG AAAAGATTT TGGAGCAA GTAAACCTAC TTCTGTCTT
 840 GCGAGCTGT AGATGATCA ACCCAGGAC TTTCGCTCA GAGAAAAAT GATTTGTGA
 900 AACTGCCAG CTCAGGATA ACCAGGACA TTCAGCTGT TTCATGGAT GTATGTTTC
 960 CACTGTGTC CTTAGGAGT GAGAAACCA TTATACTCT ACTCTCAGTA TGGATATTA

ATGATATTTTA ATATATCTGTT TACGCCCCACT AAGGCAAAAT AACCCCCAAA GAAAGCTGAC 1020
AAAAATCTGA AAAACTAAATG AGGATTTATTA AGCTAAAAC TGGAAAAATAG GAGGCTTAAA 1080
5 ATGACTGCGA GCTGATGCAAT GCTGACACTT GGGCCAC 1117

10 (2) INFORMATION FOR SEQ ID NO: 63:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

20 CCGAGCGGTG CAGGCGCCCTG GCAAGCCAGG CCGTGGAGGT TACTGTAAAG CCCGAGCTC 60
CCGCGAGCTC CCGCGGACTS CTGCCCCCTC CTTAACATGA ACCGAGTAAG TCGTGGCAGG 120
CTGAGCTGGA TTATATCTGAT GTTGCCTGCT GGCATGCTTT TAAATCTGAG GGGCTGCAATC 180
25 ATCTATGCTT GAGATGATTC TCGAAGACA GAGCTAAGA AGAAATGCC AGACAAATTC 240
TTTGGAGACA ATGAAATAT GTAACTCTTC TGGATTTAAT TATCTGAAA TACAGTTCTT 300
30 TCCCTCATGC TTATGTAAAT ATAAAAATTA ATTCATTAAT GCAAAAAAAA AAAAAAAA 360
G 361

35 (2) INFORMATION FOR SEQ ID NO: 64:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1668 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

45 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

GGCAGCAAGT CTGCGAAGCT ATTAAGCAAG GCTGTGAACA CATTGTGTAG AACAGTAGAG 60
ACTGTATAC GTGCAATGTC TTGGAGGAGT TCGGAGCTGC TGGAGATGAG AAAAGCTGCC 120
50 GAAAGAGAT GTCTGCAAAT CAAGCCACA TCGCTGCGAA CAGATTTGTG TTAAATTAAG 180
GAATCTTAC ATCTGCAAAT GCTCAKAGG ATTTGTCTTA GCTGAGAGG GAAAGAGGTG 240
55 GAAAGAAATC ACTGAAGGCC CAATTGACT GTCTTTGTG ATCGATGAT CCAAGAGTCT 300
TGGAGAAAG AATTGTGAG TCGTAAGCA GTTTGTCACT GAAATTAAG ATTCCTTGAC 360
AATTTCCCC AAGGCGCTTC GAGTGGGGCT GCTTCGATAT TCCAGCAGG TCCAGCAGA 420

GTTCAGCTG AAAAACTCA ACTGAGCCA AGACATGAAA AAAGCGTGG CCCATATGAA 480
ATACATGGA AAAGGCTCTA TGACTGGGCT GGCCTGAAA CAGATGTGTT AGAGAAATTT 540
5 TACCAAGGA GAAAGGCCCA GGGCTTTTCC ACAAGGTCG CCGAGCAGC CATTGTGTC 600
ACGAGGAGC GGGCTGAGA TGAAGTCTCC GAGTGGGCA GTTAAAGCCA GCGCAATGCT 660
ATCACTAGT ATGCTGTGAG GGTAGAAA GCGATGAG AGGATCTACA AGAGATGCTC 720
10 TCGAGGCCA GAAACAGCA TCTCTCTTAT GCGGAAGCT TCAAGCAAT GATATGATTA 780
ACTGAAAAC TGAAGAAAG CATTGTGAA GCTGTAAAG ACTCGAAGG AAGACAGAC 840
15 TCTCGAGAG GGAATGCC AAAAGCGTC GAAAGCCA CATTGCAAG CAGATTTCTG 900
TTTGAAGAG ACAAATCTTT ACGCTTACA CAAAGCTTT CCGATTCAC AAAACCTTCA 960
GAAAGCGCT TGAAGAAA ACAGATCAA TCGAATGTG AAAACCTTAT AATGTTCAG 1020
20 AACCTTCAA AGAAGAGT AAGAAATTA ACAGCGCT TGAAGAAAT GACAGCAGA 1080
ATGAAAGCC TGAAAATG CCGAGTAC AGATGAAT TGAATATGC GACATATTC 1140
25 TATCATGCT ATCGAGAT ACATGAAG CAGTCCAGAG CCGCAAGCT GAGCTATTC 1200
TTAAATCAT AATGTGTGA AGTAAACAA TCAATCTGA GAACTGCT TTGCACAGA 1260
30 ACAAAGCA GAAATTAACA CTAACTGTGA TAAATTAAG TGAAGAAA ATCTTCAGA 1320
ATTCTAAGT GAATTTACA GGTGAAGT AATTAAGT GAAAGTAT TTGTAAATTA 1380
CTGTGAGAC AACTGCTTC TGCCTCATCC TGCCTTAAGT TGCATCTCA TTGACTTGA 1440
35 GATTAAGTT TGCAGATCT TACTCTGTGA GAACATGCG CATGAGAT GCTGTTTTT 1500
TGTATGAGC TTAACTTGA TATATGTATA TGAATGTAG CATTAATCA TGAAGCATAT 1560
40 GTACTGTG AACAGTTG ATTTTATA CAATTTAA ATTCACACT TCAAGBAAA 1620
AAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 1668

45 (2) INFORMATION FOR SEQ ID NO: 65:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1353 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

55 GGTGTGACC AAGGTTGCC CCGAGCGTCC GATGTGTCG GCTGTGCTG AAGACGTTG 60
GTCTGTATG CCGCGAGGCC CAGTTTACG CTACAGCTTG TATCAAAAT GCTGTGCTT 120
60 TGGAGACAC GGCAGAAAG GAGATGAGC GGTTCGTGAA TAAATATTA GCTTCAAAC 180

217

5 GTCTCTCTC TCCCACTAT ACTATCTACA GTTGTCTCTT TCCATGCG AGTCCATCT
GCCACCGTGG CACTGTATAT GCTTTGAGTG CAGGGCTCTC TCTTTTGG AGTCCGCCC
TGTACTCC TGGAACTTT GAGTCTTAT TGGAACTTGT GAATCCCTG TGTCTGGGC
CAGCACTAT CCACTAGCT AGTTTGCAC TTGTCTTCCC TCTCATAT CATACCTGA
10 ATGGGATCCG ACACCTGATG TGGCACTAG GAAGAAGGCT GAAGATCCC CAGCTATACC
AGCTGGAGT GGTCTGCTG GTTCTACTG TGTGTCTCTC TATGGGCTG GCAAGCATGT
15 GAAGAAAGGA GGTCTCCAGC ATCATCTCC TACACATAT TACATCAC CATCTTCTG
TTTGTCTATC TTATCTCCAG CCTGGGAAA GTTCTCTCTA TTTGTTTGA TCTTTTGA
TTTTCAGATC TCTTGGAGC AGTAGATAC CTGTAGACC ATATATGAG AAAAGGCTCT
20 AGTTTTCCTC TTGTCTCTAA AGATGAGTG GCTGGAAAA CTCCCCTTTT TTGCCACAG
CTTGCCTACT CTGGCTTAG AGCAATAT TCTCTCTCCA TATGGGCTT TGAATCTGC
840 TGAGGTCAG CTTTTGGCTC CTCTCTCTG AGCAGTGA ACATGCCA GCTCTGTGC
25 TTCTGCCCCG GGGATGGCC GGGTGGGG GTGGTTGT GTGGCTTGG GTGCCACTGC
CTGTGGTTG CTGGCTAAA GGCATCTCT CTTCATGT GAGAGCCAG GCCATTAACA
1020 CCTACACAGT GTTATGAAA GAAGAGGT GGGGTGAG GGGATTAGT CTGTCCAGC
1080 TAGAGGAGA TAAAGGCG TAGTAGTTC TTGAGCGAG TGTCTTTGAG GAGAAATAT
1140 ATAGCTTTGG ACACAGGAA GATCTAGAA ATATCATTTG ACATATTA TGTCTTTTC
35 TTTTCTTGG ATTTCCAGAA AGCCTCTTA ATTATGCT TTCTCATGA AGTATGTAC
CCTTTTCTC TGAACCTGAA TTAATATCT ATTTATCT TGAATAAAA AAAAAAACC
40 TNGGGGGGGG CCCCAGACC NAATGGCCC TAT

(2) INFORMATION FOR SEQ ID NO: 66:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1011 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

55 GGGAGGAAAG CAGCATCCA GACATTTTAC AACAGTACC AGGTGTTAG TGTGACTTC
AATGACAAA GTGATCAGAT TATTTCTGT GGAATAGACA ATGATATCAA GGTCTGGAC
60 TGGCCAGAA CAGCTNACC TACACATGA GAGGCATCC AGATTCAGT ACTGGCTCA

218

5 GTTTAAGTC TGAAGCTCT TATCTTTTGT CCAATCCAT GCACATACA GTTGTGTCT
300 GGGATGTCC GCAATTTGCC CCAAGAGGA GATGTGTAAA GATATTTCAA GGAATGTCC
360 ACACCTTGA AAGAACCTT CTGAGATGTT CTGTGTACC TGTGTGAGC NAATGACAG
CTGGCTAGC GACAGGTTT GTTTATGTGT GGGATACCAC AAGCAGAGA ATATTGTATA
420 AGCTGCCGG CMTGCTGCC TCCATCAATG AAGTGGCTTT CCACTCTGAT GAGCCATCA
480 TTATCTCAC ATCGAGTAC AAGAGACTGT ATATGGAGA GATTCAGTCA AGATATGGAC
540 TGGAGACTC CAGGCGGCT TGTCTTGG ACCTCAGACT GCATTAAGTCA TCCCAATGT
600 TGGATGTCA GGTAGCACC CTGCTTCAG ATGACCTTG CTAGCAAGAA ACAGAGGGG
660 GTGGCATAT TCCAAAACC ACTCTCTGCC CATTTCCACA GATGACTNA GGCAGCTCC
720 CTGTGGCTC TAAAAACAC CTGCGAAT CTAGGACTG TTTTTTTT TCTTTTCTT
780 TTTTCTGTT TTCTAATGCA GGCCTAATGT GACAAATTTG TTGTTGGGA TTTTTTTTT
840 TTTTGTAC TGGCTGTAT GATATTTTCT TTCTGTATTT CTCTATATCA TTTGTATTA
900 AAAGCAAAAT AGATGCTTT TTACAGAGAT AAAAAAANA AAAAAAANA NAAAAAANA
960 CTGGAGGGG GGGCCGGTA CCAAAATGC GGAATGAT GGTAAACAT C
1011

(2) INFORMATION FOR SEQ ID NO: 67:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1193 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

60 GGGCGGGGG TGGCACTTC GGGCCCATCC CTGCCCCGGC GCGGTCTGTG CCCCCTGGAC
120 CTCACAGCG GTTCAGAGGG CGACTGTGTC TCAGGCCCCG GCTGACCCA GAGCGAGGC
180 TGTCCCCAGA GGAGCAGGG GTCTCTGAAA GGAAGCTGAA AAGGAACGG AAGAAAGAG
240 AAGGACGG TCTGCGGAG GAGGCGCTTG TGGCCGAGCA CCGGCTGCC AGGGCTCGG
300 GGGCGAAT GGCCTGGAC TACTCTGCA GATGGGCCA AAGGCAAG AACTGAGGT
360 TTCAGAGAC GAGGCAGAG TGGCTCTGTC TGCACATGTA TGACAGTAC AAGCTTCCG
420 ATGACACTT CTCACCTTG CTGGCTTACC TGGAGGGGCT GCAGGGCCCG GCGGAGAGC
480 TGAAGGTGCA GAAGCGGAAA GCGTGTGAG GAGCTGAT GAGGAGGGCT CTGATCCGC
540 CTTGCCGGG AGGGCCGAG GCATCCGACA GATGCTGAG CTGCTCTCT AGTGCTTCA
600 GCGCGGGGG GGGCCGCTGC CAGTGCAG GCTGCTCAG ACCACAGG GTGAGCTCC

219

TCGAGGCGGTC GGGGCGGGGT TCACACAGAG GCGAGCGGCT GACGAGAGGC TTTCAGCTCC 660
TCCGATGATG GGGGCGGGGA TCACACAGAG CAGAGCGTCG CAGAGGCCCC TTCCCTCTTC 720
CAGAGCGCTCC TTGGCGGGAT AACCTGTGAC AGTAAATGCA GATTGAGTGC GTTCAGCGCA 780
GAGGTGGAAT GCTGTGGAAT CAGCCGAGAC CTTGCGCTTG GAGGAGACCT CAGCGCCGAG 840
GAATCTGCTT TCGAGCGGAA TGTCTATTTT TCTACCGGAA ATATTTTACA GATTGCGGCA 900
TGCCTGCTCC TCCCGCGAGC TCGAACCTTG GAGCTTCGCG CTGATTCGCG ATCCCGCTGC 960
GTGGCGCGGA TTCTGTGTCG CTTGCGCTTG TCCATGAGAG GCGCTGCGCT TGCGCTGTCT 1020
TCCCTTGACC CCAACAGAGG TCAATTCGAG TCAATGAGAG CCGCTGTGAG GAGCTGTGAG 1080
AGTCGAGTCA GGTACCTGTG CAGAGAGCGC CTCGTGCTCT GCAATTTGAA TGAAGCCGCA 1140
CCGACAGACA AAAAAAAAAA AAAAAAAAAA NAGGGGGGGC CCGAGACCA ATT 1193

25 (2) INFORMATION FOR SEQ ID NO: 68:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 560 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

GAATTCGACA CGAGTTGACA CAGATGACA AATGCAATTC TCAAGTAGA TTGCAATGCA 60
AAATGTGGA AACATGAG CAGTGCACA TAGCATGCAAT GCTGTGCTG ACCTGCAGCA 120
TATTTCTCC TTCTCTCCTT TCTCCCTCAT TTATTCATTC ATTACATGAT TCAATTCATCC 180
CAATTAATAA ATTATATGTA TGTTTGTGTC AAAGACCTT ACTGAGGCT GCGGGGTACA 240
AAAGATATTC AAAGGCTTGC GAGCTTGAGC NAGTCTGAG TAGTAGTCT AGATTTGTGT 300
GAATTCGACA GACTTACTCC AGGCTCTTGG TGAAGTGTGC TTGCAATTA TCTCTTAGGC 360
TAGAGCAATT ACCTTTCAT TTATCAATCT TTGCTGATGC TAGAGACAGA TTCCAAAGTG 420
CCCTCTTATT AATTTTGTGA TTAAATGACA AGTATATGCA AGAATAGGCG ATTGTAGCT 480
CAATTCCTTT TCTATATTTA TCTTTTCAA CAGTAAATTA TCAAGTAGAT GAAAAAGGCG 540
CGAATAAAAA AAAAAAAAAA 560

(2) INFORMATION FOR SEQ ID NO: 69:

(1) SEQUENCE CHARACTERISTICS:

220

(A) LENGTH: 1657 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

CGAGATAGAC CCGGCGCGGG CACTTCCTGT GAGAGCGCGCA GCGGGTCCGG GCGCGCAGCG 60
CGAGAGCGCA GCGAGCGAGAC GAGCGAGAGC AGCGAGACTT CCGGCGCTTG CCAATGGGCA 120
GAGAGCTTG ATGGGCGAGC CCGAGAGCTT GCGCGACAG TTCTCTCTG TCACAAAGCA 180
GTACCTTCCC GAGTGTGGGC GCGTGTGTCT GATCAGACAG TTCTGTAGAG AGCGCAATCG 240
TATGTGTTC CAGTGTAGAGC AGCAGCGGGA CTACATGAC ACCACATGGA ACTGCGGCTA 300
CTGTGTGGCC TGTGTCTGTC TCTTCTTCAA CTTCCTGGAA CAGTACATG CTGCGTCTCG 360
GTATTGAGCA GAGACTTCCT GCGATAGCCC TGCCTTGGGG TCTTTGGAAT CAGAGCTG 420
CAGAGATTC CCAACAGAT TTATGTGAC TTGAGTTT TGAATGAGAA CTTGCGCTCG 480
GAGAGAGGCC TGTGTGCTCT CCAACAGAAA TCCCTGTCTG AAGGAGAGAG CAGTTTCCG 540
GCGTCCCCA CCAATGCTGA GAGCTCCCCC AAACAGTACA TGCAGCTGCG AAGCAGGCTC 600
TTGTGTATTC TGAATTCAT GAGCTTCTT CACTTGAGC CCACTTCCTT TTCTATGTC 660
CAGAGATTC TGGGCGAGC CTCGTGATAT TTATGTGACC ATGGGTTTGA AAACAGACT 720
GCTGTGTTG ACTTGTGTC TGTGTCTT TGCATGAC GATATTTCA AGGCTGTG 780
GAGCTTCCA GTTCACAGC CCAATGATTA CTTCCTGAAA TAGAGCTCT TCCAGCAT 840
GTGAGTAT TGGGGCTTCC TCCGTGTGAT GCGCTTGGGC CTTGGGGTGC TCTCATGGA 900
TGAAGAGAG AAGGATGTG ATACATGACA GATCCCTTAC TGCCTGCTTA AAGCCGTGCG 960
CCGTGAGGA CTGTTGCGGG GTGATATCAA CAAAGTCCG AGCTTTTAG TATCTCTTTC 1020
CTTCCCTTCT CTTGTGTAAA GCGACAGTGT TTGTGAGAC TTATATTTCA GAGACACTG 1080
AGATATGAT GCTTGTAGAC ATGGGTCTTC TTCTCTGTG ATCTATGAG TCTCATGTC 1140
ATTAAGACT GCTTGTGCA GTTGTGAC CTTACAGGG GTGTCTTGG GTCCAGACT 1200
CGCTCTCTT TAGTGCAGTA TGAAGCCCC CATTCAATGA CTTTGGCATA GTACAGTCT 1260
CTCTGTGATC AAGCTTGGTT GAGTGAATGG TGAAGATGAG GGTGAGACAA AGAGAGCCAC 1320
GTAGAGATC AGCAGCATTT CTGACAGAGC AGGCGCTCG TCTATGAGG TGTCTCTT 1380
TCTCTTGGCC GTGGTGTGAG TGAAGCTTGA TTGCGGAGAA TCCCTTTTCA GGAAGGCGAG 1440
GATAGTGGG ATCTACAT TGAATCTGAG AAAACATTTT CTAAGATTTT TTCTCTTAT 1500
GTGGAGAAC GATCTAATTC TCAATTTAG GTGATATTTA TATCTTAAAT GTCTTTGAAA 1560
AGCTTTGAT TTGTGAGAC AGATCAAAAT AATTAATGCG GTTGTGTGA AAAAAAAAAA 1620

221

222

AAAAAAAACTC GGGGGGGGC CCGTACCA AATGGC

1657

5

(2) INFORMATION FOR SEQ ID NO: 70:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 711 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(A) LENGTH: 711 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

GGCAGAGCG AAGACCTCT TGGACCTG CCGGATTC AGACTAGGT AGATCTGCG 60
CATACCTCT ACCGTGACA CAGGAGCC CTGGGCTGA TGGAGAGA TCAGGTATC 120
CCAGGAGT AGGGCTACC TTGAGGGAT GATGACCTC CCGCACTCC AGTGGACTC 180
TGGAAATAT AGGAACACT GAGTGGAG AGATTTCAG CTGCGGAGA GGAATCTTC 240
CCTTCAGGC CAGCACTGC CTTTGGGAA TGTGCGGGG TCTCTCTTT CTCCTGCTG 300
TTTAGGTGG TACAGATTC CCCTTCAG TGGGGGAG CTGTCGGA CARACTCATC 360
TCAGCTTTC CTTGGGAG GATCGGGGC AGCAGCTCA GCGAAGAC CAGGATCTG 420
AGCAGAGCG CCTTGAAGG ACACAGGGC TGTGCGCGG CAGTGGGCC CACCTCTCT 480
GGRAGTGG CAGCTCTTC CAGGCTTTC TGAAGAGGA GAGCAGGCT TATGCTTAG 540
CTTCATAGT TCTGCTGGC TGGGTGGAC CAGAGAGCC TGGGGCTGG GTGCGCTAG 600
TGTGTTAAA GTGAGCAAT CCTTCAGC TCTTGCCA TGTCTGAG GGCAGCTTG 660
GCTTCTCT TATTAATGT CTTTATTT CAAAAAAA AAAAAAAAAA T 711

(2) INFORMATION FOR SEQ ID NO: 71:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 935 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(A) LENGTH: 935 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

GGCAGAGGT GAAGGCAGC TAAACCCAA GTGGAGAT GAAAGATG GTTGTCCA 60
TAAGTTTAT GCTCAGATA TGAAGAGC CATATGAT AGTGAACC TCCATAGTT 120
GATAGAAA CCACAGGA AGATCTCTT CTGGAGAG CAGCAGCT CCGTAGAG 180
CGGCCAGC GCGGGGCGG AGGTGGCT TTGTGTGGA GTGGAGAT TGTCTTCC 240

CAGACATCT CCTCTGAGC AGAAGGCAG TATGTTTGA GACTTTATG TATGAGAGC 300

GACAGTCAA CGCTGCCCG GAGTCCCC TTGTTCGAA ATACTTTGA AAGAGAAC 360

CTTGGTATA AGCAGTATG CAGTCTTCC CTGGCTGAG TCATGGGCG CAGCTCCCTG 420

GACTTGAGC TGAATCTCA GGGTTCGAA ACACGGAGA GGCAGCTGA TGAAGAGCTC 480

TCCGCTTCC GTGAGCTGG GCAAGGTTN GGAGAGGCC CAGCTCCCTG GCCAGACTGA 540

CCTCCAGCC TGGTGTCTC GGGAGAGCG GCTCGTGGC CTGCTGGGG AGCGGAGCG 600

CAGCAGAGC AGACCAACT TGAATACCT CATGAGCG CGGCTGAGAA GATGCTGAG 660

AAGGCTTCA AGGATATCT CCGCTGGCT GCGAGGCC CAGAGAGCC ATCCAGTGC 720

AGACTTTAG GAGAGATA GCATCTTCA CAGGCCAG GATCAACATA CCTCTCTCC 780

CAGCGAGCA CTTCTGATG AGTGCATGT GCACATGAG TATTTATCCA CCGTGTAT 840

TTTCATGAG TTCTTAGCT ACCTGAATTT GTCTTAAA TATTTGTGA AGCTATTA 900

TATACATTT TTGTAAAA AAAAAAAAAA AACT 935

(2) INFORMATION FOR SEQ ID NO: 72:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(A) LENGTH: 504 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

GGAGGGGGA GGGGTGGG ACCCGGGG GAGCGGAGC GAGTATGTC CTTGAGCTC 60
GGCTGGCTC TTTCGCTGC GTTGGAGCC GCTTTTCAG AAGCGGCTC GCACGGACTG 120
CTGGAGTGG TGGATGCTT CATCGGCTG GTGGTGTGA CATTGAGCC CGGTATGAG 180
AGTTCCCCA GCTGACAGA TCCGAGTGT TCGAGCGA GTTCTTACG GGAATCAT 240
GGTTCTGAT TCTTGCGCC TTTCGCTAG ACTCAGAGA GGTCTGGGT CACTTTCCT 300
ATCTGATCC TTCCAGTGG ACAGTGGAG AATTAGTAT CCTCTCTGAT GATGAGCT 360
GAGGTGTAG ACTAGCCTC ACTCTGTACA AGAGCAGGT GAGATTTCA AGGATATGG 420
ACTTATTT GCACATTA GTTACAAAT AAGTGGCTT GGTCAAGAT GAAAAAAA 480
AAAAAAAAA GGGGGGGGC CCN 504

(2) INFORMATION FOR SEQ ID NO: 73:

223

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 620 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

10 GAATTCGCGA CGAGGAGGAG GGGAGGCGGG GTAACTTTGG TGGAAACTC TGTATTCTCC
ATTATTACTT TGCAGCAAT AGTCAGCAAT CCAAGATGGA TGTCTCTTT GAGCCACTT
TTGCTGTGCC ACTTATCTTG GAGCAAGCAAT ATGAGAGTGA AGAAAGACTG GAGAGCAATG
AATATATCTA GGTGTCTGAT TATTATACAG TCATCCCCAG TTATGATGAC TTATGTCGAG
ATTTCACAT TGAATTACTC ATATTGAGT CAGAGGACAG GCTGAAACAG TTGATTTAAG
ACATATACAG AGCAATAGAG ACTACACTTA GTCTTGAAAC AGCACTGTCA GACCATCCGA
AGCTGTAC TCAGAAACA GTACAGAG AGCTGAGAG TCCAGATG AGCATGCGG
TGTCAAGTT GGAAGTCTT ATTCCCTCC TCTGTCTGTG TGTCTTGTG CAGGTGGGGA
TGTATTTCAT GTTAAAGGTG GAAAGAGGCT GCTATGACTC TTGAGATGG AGTCTGGCA
GAGCAATTTG GAAAGTAAA TAAATATATA GTGAAATATA AAAAAATCA AAAAACTGA
30 GGGGGGGCCC GATACCCCAAT 620

(2) INFORMATION FOR SEQ ID NO: 74:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 581 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

45 AAGAGGTGG TGTAAAGTT ATGTTTGTAA ACTGAAATCT ATCTTAATC CAATAAGAC
TCCAGAGTAA TTCAATTTTG TACATTAAG ATCCCAAGT TTATTTTGA AATATCTAT
TTTACACAT TAAATAATTA CAGGCAATG AAGAGATTC TACGTGACAT CCAAGCTCT
TTAGCTTTTG GTGTGTGCA CCGGTATGTC AGCTCTCTC TCTCTTCTG CACTGTCTGA
CAAGCAATG CCTTCACAG CCGGCAAGT GCTTCTCTT CATGTACAT TTGAGCTTCT
GCTCATGCTG CCTCCCTCC CCTCCCTCC CTCCTAACCC GGGCTTTT GTTCTCTCAT
GAGATGCTTC CATGGGTGG CCTCCCCAG CCAAGCAATA AATAGTGGTT TCCCTTTCG
TTCTGTAGCC CTTCAGACA TCTCTGTATT ACAGTATGG TTGACTTACT TCCCTCTCC
60 480

224

CGGTAAAG CATTAATCC TTAAAGACAG GTAGCTTTG TATATCTTC GTTCTCTCA 540
ATGACCAATA GACATTTTAA CATGTACAA ACAATTTGA A 581

(2) INFORMATION FOR SEQ ID NO: 75:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1843 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

10 AAACCAACN CCTCCCGATC CCAAAAAGA AGCCAGCCC AATTCGAG CCAGCATGTA 60
GCTCCGAC AAGCCCTCA AGAGCCGAG NCGACAGAC AGCCCCAG AGGCCCTGCA 120
AAGAACTCC CTGGCGGCC AAGCGGCGAG CTGTGTGTG GAGAACTCA GCGACGAGA 180
GCGAGACAG CATGCAAT TATGTCCG AGACCAAGC CAGGCTGTGA GACATGAG 240
25 GAGAAAGA ACGATTTTAA ATCATTTAAA ACACAAAAC TAACTGGTAA CCGACAGAG
TTTCTCAC CTTCGTATG GTTATCTGT CTAGAGAGCC TGAAGCACT TTCAATTTGA 360
GCAATGAG GGTCAAT TTGACTTTT TGGTCCCTC CCACTTTAG GTTATGAGA 420
30 TTTTACTAC AAAAAAATC AACAAAATC AGCAACTAG AAAAACTTT TTTCTCTCT 480
GCTGCGGTG GTGACTAGA TTAGTGAAG TGGCACTTC CCGCGCAGC CTGCATCTG 540
CGATCTTTT ACTGTCTTA TCTGATAGA ACTGACATA GCTGTTTAC AAGATGACA 600
40 CAGTCAGAG GCAAGCTTG GCACTGCA TGTCTCTCT TTTCCCACT ATCCCGGCTC 660
TGACTTGAAT TTCAATCTT ATGTTTCT CTTTTCCCT CAGAGCTCAC ACAATGTCTA 720
CGATTTGAC AAGCGCTTT CTGGTGTGA GCTCTCTG CGATTGAGG CCGAAGAGAC 780
45 AAGAGATGG ACATGCTCC CCTCCCTCC CCGGCAAGT GCTACAGAC AACCTAGAGC 840
GAGACAGAC ACACCAAT GAGGCGCTT CACTGGAAG TGCCTCCCA GCGCTGGGCA 900
GTGTAGGCA GATGCTATC AGCTGTAGC AAGATGAGA AATTTTATG TTGGCGGATG 960
50 GAAATGAGC GAGGCAAG TTATCAACT CAGACTCT GCTCTTAT TTTCTCAC 1020
CCTCTTCC TTGACAAA TCTAGATTC CCAATGCTT TCCAGGGGCG GTTCAGTCT 1080
CAGCGCGCC TGTGTGCTT GCGCAAGGG CAGGCGGCG TGTCTGTATG TATGTGTACA 1140
55 TATGACATA GACTTGAAG TTATATTTA ACAAAAGCC ATCTGTAC CCAATGCAAC 1200
CAGGCGCCG GCGGCTGAG TCTCGGCGCA CTTCGAGGA GCGGAGGTG TGAATGAGAT 1260
60 AATTTTAC ATGTACTATA TGTAGTGG TGTACAGTG TGTGTTAAA TATATCTCT 1320

225

5 GTGTGTAAGC AGCCCTTTT TTTTGTGTC TCACACCCC TCCCCCGCC CCGCACTCT 1380
AAGGCCCCAT CTGCCCAGCC TCTGAGTTT CTGTCTAT TTTTTTTAA CCCCATTTT 1440
CCTTCTCTCT CTCTGCCCC GCATCCAC TCCAGGGTG TCACAGCCC TGAGTGCAT 1500
TGGCCCCGGC CTGAGGGGG GGTAGGGGA GGGCARGCT SAGCCCCGA GCCAGCTCAG 1560
10 TACTGAGGG CCTGCTCTAT CGTGTGTATG GCGTCTCTG GCATCCGGA CATCTCTTG 1620
GTGGCGTTG CTNGCAGGG AGCCCCCCC CGTCCCGAG TGACCCAGG GTCTGCTCG 1680
GGGCCATTT CCAGCTTGGC GCGCTCTGT GACCTTGGG AAGTCACTG ACCTCTGT 1740
15 GCTCACTT CTCTCTCTT AAAAGGGGA CAGTCCCTGC CCGTCCCTAC CTCACAGGA 1800
TGTCTGAGA ATAATGAGG TAAGTGTAA AAAAAAAA AAT 1843

20

25 (1) INFORMATION FOR SEQ ID NO: 76:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

TCGACCCAGC GGTCCGCTC CCGAGCCCT GCCACCATC GTGACTTGG GTCTGTCCG 60
35 GTTGAGACAC GCGTGGCTG CCAAGCACC GCGACTCGG GAGTATGCG CATGCCATC 120
ACAGCCCTTC ATGAGAGGG TTTTCACTT GGTACAGGC ACGGCATCG CTTTGGCTT 180
40 GCAGATCTTC ATTGAGAGA AGTTTCAATA CCTTTGCGAG TGGAGCCTTC TAGTGGCGT 240
GGTTGCAGGC TCTGTGTCA GCTACGGGT GACGAGACTG GATCGGAGA AATCCACAA 300
45 CTTCTGGCTC TTCTTGAGA CCGGCGAGCT CCCAAGAC AGGAGCAGAG ATCAGAGAG 360
CTAGAGAGC TCAGCAGGG GCACAGAGA TTGGGGGAG GAGGAGTCTG GACACAGCC 420
TTCAATGCCC CTGACCCAG GCGACCTTC CCCACACCT AGGCTACCCC AGTGTATTC 480
50 TCTGTCCGA TGTGTGCGA GCGCTGACA ACAGTGCAG ATGGTGTGT CCCCACCTG 540
GAGCTGCCC AGRAGTTTG AGCGAAGG GCTCTCCCTG GGTGTGTGT TCTCTCTAG 600
GGTATGGA TGATGTCTT GACTTCCAG CAGAGAGGT GTGTCTGGG GCCACACCT 660
55 ATGGGACAG GGTGGAAG GCGCTGACA CTCTGTCA TTCTTTCTAG CCGCTGCATC 720
TCCACACAGT CCAGGTGAC AGCTGTGCT AGGGGCTG GGTAAATA TGGCTTATC 780
60 TTCTCTCAC CCAGTTTTC ACCTGACAG GTCAAAAAA ATTCAGAGG GTAAAGTAT 840

226

GACAGGTAC ATGAACCTT TATTACCTA CAGTTGATAT ATGAGATCA CATCAAGTT 900
ACATCTAG CAGTATCAG GAGTTTCCA GCGTGACC CCAAGTTAG AGCTTGCAT 960
5 CAGCCCCGTA GGCAGGTGG ACACACAC AGCTCTCTG TATGGGGTC TGCTCTGTA 1020
GCACTTCCA TGTAGGSCA GAGCAAGG GCGCAGCTG GCGAGGCTT GCTCTCTGG 1080
NAGAGAGGG ACTTGTGGG CAGGCACTT GCTATCATT CCCCATCAT CTATTAGCA 1140
10 AAGTCACTCC CCGAGGGAG AGCTAGCCG TTGTAGCCGT GTCTGTGTG AGGAGAGCT 1200
TCTGAGTGG CAGGCTACA CACAGCCCC AGCCCCAAGA GAGAGAGAG GTGAGACCA 1260
GACGACCTT CCACAGTCC ATCATGTTA CAGCTGGCTT CCCCCAGCA CCGAAGACC 1320
15 ACAGCTTGG CCGTCTGCC CCGACCCAG CTCAGTCCC AAGCTCACC TTGCCAGAA 1380
TTGAAAGAA GTTATGAGT ACTATTTGC CTCAGAGTAA CAGGAGCTC AAGTTAAGT 1440
20 G 1441

25 (2) INFORMATION FOR SEQ ID NO: 77:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 910 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

GGCAGAGCTG GCTTCGACT CCGTATGTC ACTAACATA TGTGGAGCC AGCGAGGCG 60
120 AACAAATGC TTAGGTGAG ACCCAAGGT CGTGGCAGC GGTTCGGGT GTGGGTGTG
180 ATCGGGGCC TGGAGGCC CTCTATATC CCGGGCAGG ACCTGAGCC CCGTACCTT
240 CGAGCTGTC GCAGTACAA GCCCCGCCG AGCGAATGA ACCGGGCTT GCACGACCG
300 ACCCGGACT ACATGAACCT GCTGGGATG ATCTTACGA TGTGGGGCT CATCTTAAG
360 CTGAAGTGT GTGCTGGGT CCGTGTCTAC TCGTCTTCA TCAGCTTTGC CAGCTCTGG
420 AGCTCGAGC ACAGAGCA ATGATGAGT AGCTTCATG GAGACTTCC CTACAGACA
480 AGTGACTCTT GAGTAAGGG TGGGGGACC CAGGCTGGC CATCTTAGAC TGACACTCT
540 CTCTGTCTT CATGCTGTC ATCTGTGCC TGGTATGTC CTATCTGAG AATCTCAGC
600 CATTAGGCC CCGATGTGA TACAGGCTA GAGGGTCAG ATTTTGACC CTGTCTATCC
660 ACTAGGCGT GCGTTGGCT GCTAAACTG CTGCTTTCG CTGCCATCT GACTTCCCT
720 GAATGAGCC GTCTGGTGC CCGCAGTGG ATAGAGGGA CCTGGGCTT TCTTAGGGA
780 CAGCTTAGC TTACCCCTC TCGCTCCCTT CCGCTGCCG CTCTGGGG AGATCTCTC

CAATCTTCTA GGGGATTTCA TTTCCTTTCT CATTGAAACC TATTGTTAAT AAGATTTTTC 840
ACTCTGAAAA AAAAAAAAAA AAAAAAAAAA TGTGTGAGGG GCGCGAACC GAAATCCCG 900
GATATGTAGT 910

10 (2) INFORMATION FOR SEQ ID NO: 78:

15 (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

TCGACCCGAG CTTCCGAGCG GCGATGATG GCGGCTGTG ATGCGGACGT GAACTTAGGC 60
ACCTTGAGGA GTGCGACGGA GACGACGAC GACGCGGCGA GCGAGATGCC AGCGAAGCCG 120
GAGCGGCGAG CGAAGGAGGG AGCTGAGGCG GCGGCGGCGT TGGCGCTCTT GACCGGAGGC 180
GGGGAATATG TGTGTAGACT GCGCGTGGTG GCTCTGTGTC TGTCTGAGGC CTACCGGCTG 240
TGGGTGCGCT GGGGCGGCGG GGGTCTGAGG GCGCGGAGCG GAGCGAGGCC 300
GCGACCTCTC TGCCTGCAAT GAGAGAGCGG GACTTCAGCT TGGAGCAGCT GCGCGAGTAC 360
GACGCTCC GCAACCCCGG CATTCTCTC GCGGTCAATG GGAAGCTTT CGAGCTGACC 420
AAGCGACAGA AGTTCTACGG CCGCGCGAGT CCAATGAGA TATTTCCTGG TAGGGAATCC 480
TCCAGAGAGC TGGCAGANT TTGCTTAGAT AAGATGAC TTAGAGATGA AATGATGAT 540
CTCTCAATTT TGAATCAAGT ACAAATGAGG AGTGTGAGG AATGGGAAT GCGATTGAAA 600
GAAAAATATG AATTATGAGG CAGACTCTTA AAGCCAGAGG AAGAACCATC AGAATATACA 660
GATGAGAGAG ATACGAGAGA TCGAGATAAA CAGAGTTGAA CTTTGTAAAC AACGAAATC 720
AGGGGCTTTC AGAATCGAAA TTCTTAGTCC CTTCGACAGA CTGTCCGAGG TCTTGGGTT 780
TGAATCACT GCTCGGAAAA AGATTCAACA AATTGTGAC AGAATAAAT AATTCACATA 840
TGAAGATTTG AATTACATGA CATTATTAAT GCTGCGAACC TCAATTTGTG CAATGTGTTG 900
TAAATGTGAG TGGGCGCTTCA TCAATCTGAA AAGAGAGAGA CAGGATTTT TTTAAAGAGC 960
AAGAAATGCA GAAATTAAT TCTTCTCTC CTTTTCTCT TCTTCTCTT CTCTTCTCTC 1020
TTTCTTCTCT TTTAAATAT ATTGAGACA ACCGATATG TATTGTGAC TCAAGTGAC 1080
AGATCTCTTC AAGAAATATC AAGGAGTCC TGTGTCAAT ACTGTGTT TATTATGACA 1140
TGGGTGAGGG AGCGAATCTG ATCGAGGAGG GTGAGGATAC ACATCAATTT GAGTGTCTCA 1200

GCGTACGAAA ACATTAATAT GTAAATGCC AAGCTTTCT TTTCGCTTT GTACAGGAAA 1260
AGAAAAATAT CTTTAAAGAG AATCTTTGG AATTAGAGG AAGAAATTC AGGTGGGTTT 1320
AATCGAGAGC TAACTTCCCA ACGAAGAAAT CATTGAAAC CAGTTTAT CCGTCTCTT 1380
TCCCTCCCTT TCCATTAATC AATCAATAT TAAATGCGC TTAATCACT TAACTAGAC 1440
TGAATTAAT TTAGGAAA GCGCTTAAA TGAATCAGA AATCCACGA AGCAGATTA 1500
AGACTAGAT TGAATATTC TGTGACAT AAAACCTGA TATCATCTG TGTATATGA 1560
ATGTAAAGAG AATATATGAG TGTATATCC CATATATGTA AATATACAAA ACTCAATTAG 1620
CAATGTATAG GCGAATGCA TTCCCGCATG CTTTTCTGTT TTCAAAAA TTGAAAAACA 1680
AATCAATCT TATCCCAAC AGCTGCTTAA TTTTAGAGT CTGACCTCC ACATCTCACT 1740
GCTGTGAGTG CATTGGAGTG TGAATGGGT GTCAATATGG AATGTCTGTA ATGTGTAGG 1800
CGTGGAGAG GACTTTTCT GAGATATCTG TAAATACAG TACCATTTTA ATTAAGCATG 1860
TACATTAATC CAATATATG TTAGATTGGA CTTTATATAC AGAATGTAA GCGAGTCAT 1920
TATATATGAG TTGTATAGAT GTGCAATTGA TTCAAGATTA GAAAAATCT TCGAATATGA 1980
AAGCGAGAC AAGTTAATAC AGTGTATAC AATTGTACAC ATTGAGCATC ACTTTAGAAA 2040
GAATATGAC TTGTGAGAAA TTCTCAGAAA ATCCAAATTT ATTCAAGTGA GAATGTGAT 2100
AATTAATGTA CATTCTTTTA CTTCATTAAT TGAATGACAC TGAATATCT AGCAATTAG 2160
CACTCAAGT GATTATGAAA CAGAAAGAG TAAATGAAAT TTAGCATGAT AGAATCTAAA 2220
TTTATTAATTT AATTAGTGA AAAACCCCTT ACTAAAGCC CTTCAATATAT GAATTAATTT 2280
ATTCAATAT GACTACTTAG GTTCCGAGCT GGGAGCAAGT TCACTTAAAA AGCGAATTT 2340
ATTTAAGAG TCAACAGTTA AGACTTCTGC TTGTATATGA CATCGAGAG CCGATCAACA 2400
AAGGGAGCT TTATACATGA ACAAATGCT AAGATATGTA AATATACATA TTCTATTCAG 2460
TGTGAGATG TTTGTATGAA AGTATCTTC AGCGAATCT TTGCTGAGGA GTGCTGTGCG 2520
AGTGTGATG AATGCTTTGT GTTTTATAT AAAATTTTTT CTAAAGAAA AATGTATAAA 2580
GTACATGCC TGTGTATGAA ACTGATATCT AATATATGA ATCAATCAAG CCGAATATCT 2640
AGTATTAAC TGTACTGTG AATGAGAAA CCGTAATAT TCAATCGAGA AAAATATGTC 2700
GATCTGTAAA GAAAAATGAG TAAATGTGAT TTGCACTTGG AAAATACAGG TGTCTGTGAG 2760
AATATCTCTC AACTTC 2776

55 (2) INFORMATION FOR SEQ ID NO: 79:

60 (1) SEQUENCE CHARACTERISTICS:

229

(A) LENGTH: 1525 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

CGCTCTCTGA TAACTATGCG ATCCCCGGG CTTGCAAGGAA TTGCGCAGCG AGCTAGCGCG 60
CGCGCTGGCT CTTGCTGACA CTTGAGGCT CGTGCAGGCT GAGCGCCAGC CAAGACATCA 120
GCATACAGCA CCAAGTGGG GCGCAGAGC TGGCCGTGTT CCGGAACTTG TCCCTGCTGG 180
TGCTGGGTGT CCGCGCGGTG TTCTGCTGC TATTTCACTT GGGCAGCCCG GAGAGGGGCC 240
GGCCGCATGC GGAAGAGCA GCGAGAGCA CCCCCCTGTT GGCCTCTGCC ACGGCCGACG 300
CCCTGCTGCT CTCAGAGCAC TGCGTCCGCG AGCGGGCTTT CTACCAAGTG GGCATACTGT 360
ACATGACAC CAGGCTCATC GTGAACCTGT CCGAGACTTA CATTGGCAGT TACTCTCACT 420
ACTGCTTCCA CTCGCGCAG AAGTTCAATC GAGCACTTCC CTTGGTGATG TACTCTCAGG 480
GCTTCTTGTC CTCCTTCTTC ATGAGAGCCA TCACACAGTG CATTGGAGAG AACATGACT 540
ACTTCTCAGG CTCCTCTGTT ATCTGAGCT TTGCGCGCTG GTTGGCGCTG CCGGAGGAGC 600
TGCGTGTGCG CTGTGAGCCA GCGGCTGTGC TGCTGGGTGC TGCGTGTGCC ACCATCTCG 660
TCACCTGGCT GGCATGAGG GCGGAGCTCA TGCGTCCCCA CAGCAAGAGC GCACTTCTGT 720
GTACGGCTCC ATGAGCTTCT TGAATAGGT GGCACATGGG CTGGCAGTCA TGGCATTCCA 780
GAGCTCTGAC CTGTGCCCCT CAGAGCTCTG CTCGAGGGCC TGCGTGTAGT TTTTACAGTG 840
GCGCATGGTG GCTGTGAGCG GCGGGGTGGG GTTGGCGGCT GCGCTGTGTC TCTGTAGGCT 900
CCTGCTGTGG CCGAGCCGCG TGCGAGCGTG GAGCCGTGAT GCGCGGCCCT GACTCTCTGAC 960
AGCCTCTCTC ACTGTGTCAA GGGAACTGTG GGGAGCGAGG AGGATGCCCC CCAAGGGCGTT 1020
GGGGAAGAGC CCGCACTGCC CTCCTACTTT CTCCTGAGCC CCACTCTCCA TCCCTCAGCCA 1080
GCTCCCGCGG GTGGGGTGGG GTGAGGGCAG CAGGGATGCC CCGCAGGAGC TTGCAAGGAC 1140
CCCTCGGTTT TTGAGGGTGT CCAATTTCTA ACTCTAATCC ATCCAGAGCC TCTGAGAGAT 1200
TTGGGGTGGC CTTCTTGGCA GGGAGAGCA AGTGGGAATC CAGAGAGGCT CTGGGGGAGC 1260
CCTTAACCTG AGCTCATGTC AGTTACCCC TCACCTCCAG CTTGGGGGTC TCCAGAGACT 1320
GCCAGGCCCC CTTGAGGAG CTTGGAGCTT GAGGAGAGCA GCCACGGGCT GGTGGGCTGG 1380
GCTTGAAGCC CACCTGTGTT GGCAGAGGG CTGCGCGGCA GGCCTGGGTG ACTCTGCTGG 1440
CAGCAATATA AGAGTAGAG CCAAAAAAAA AAAAAAATAA AAAAAAATAA AAAAAAATAA 1500
AAAAAAAAA AATCCAGG TCCTGC 1525

60

230

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1563 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

AATTCGGCAC GAGACAGAAA CTTGCGGAAA ATGCTAGCGA TGGCGGCTGG GCGGAGTGGG 60
TGTCTGGTGC CGGCGTTTGG GCTACGGTTG TTGTTGGCCA CTGTGCTTCA AGCGGTGTCT 120
GCTTTTGGGG CAGAGTTTTC ATCGAGGCA TCGAGAGAT TAGGCTTTTC TAGCAACTTG 180
CTTTGCAAGT CTTGTGATCT TCTGGAGAG TTCACCTGC TTCACTGGA TCCCTGATTC 240
AGAGATGCT GTCAAGAGGA AGCAGATTT GAAACAAAA AGCTGTATGC AGGAGTATT 300
CTTGAAGTTT GTGATGAAA ATTGGGAGG TTGCTCTCAG TCCAGGCTTT TGTTAGAGT 360
GATTAACCCA AACTGTTTCAG AGGACTGCAA ATCAAGTATG TCGTGTGTTT AGACCTGTA 420
TTAAGCTTT TGGAGACAA TGGGACATTT GCTGAGAGAC TGAAGCAATCT CAATGCGAAC 480
ACAGACAGTG TAGAAGATTT CTTGAGTGAA AAGTTGGGAC GCATATAAAT CTTGCTTAAA 540
TTTTGTCTTA TCTTTTGT TTACTTATCAA ATGAATATTT ACAGCAGCTA GAATATATTT 600
TAGTTTGTCT TCTTTCATTT GATCAGTCTT TTACTTGAGG CATTAATAT CTATTAATAT 660
CCTGAAATGG CAGTATATGTC CATGATATCT AAGGAGTTGG CAAGCTTAA ACACCCATTT 720
TTTTATAAAT GTCCATCTTC CTGCAATTTGT TGATACCACT AACAAAATGC TTTGTACAG 780
ACTTGGCGTT AATATGCAA ATGATAGTTT CTGATATATG GTCCAGTTTT ACAGACAGCA 840
GATTTCTTAA TTGAGAGGT TAAACAGACA GATGATTAAT ATGCGTCAATG TGTGTGTGTC 900
TCTTTGAAAG GATGACAGC AGACTACAAA GCATTAAGCA TATACTGAGC CTCACAGAT 960
TCCCTGCTCC TCAGAGTCTC TCTATTTTTT GTATACCCA GCTTTCTTTT TAATACAAAT 1020
GTATATTAAT GTTTACAGT AATGCACTGC ATTAATAACT TGTAGCTTCA TTAATGTAAA 1080
ACATATCAA GATCTACAG TAAGAGTGAA ACATTCACAA AGATTTGCTT TAATGAGAC 1140
TACACAGAAA ACCTTTCTAG GGAATTTGTT GGAATCAGTA CATACTTGGC AATTTTTTGA 1200
GTTTATCAT CTTACAGAAA AGTCAATTTA AAGTGATCA TTTGTAGAGC CAATATATTA 1260
ATTAATAAGT TCAAAATCT ATCTGAATTT GGAATCTTC TGGTTTGTTC TTTCAATGTT 1320
AAAAATGAT TTTTTCATG CATTTTTTTC ATGTAGGCC TTTTITTAGC CAATATGTA 1380
AATGCTGT ANATTTTAA ACTTATACA TCTTATTTT GTTATATG CTTATATTT 1440

231

GTCTGATTTT ATTTCACAAA GTTTTTCAT TTATGACAC ATTTCATTC GTATATTTT 1500
TAAAGAAAT CTCCTGATAT AGAATTTTAA TATTAATAAT GATTTTCTT TCCCTAAAAA 1550
AAA 1563

10 (2) INFORMATION FOR SEQ ID NO: 81:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(*1) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

TCGACCTTCG CAGATGAGG GTTGGAGCAC TCGACCCCC GGCCTGCAC GGGCCGCGACG 60
CTGGCCCCCC TGGGGCTGCT GGCCTGCCTG GGCCTGAGCC ACGAATTCGG CGGTCTGGTG 120
CTGACCCCCG TGGGCAACGA GTTACGCTTC CCCCAGACGA GACGCTGAT GGCCTGAGCT 180
GGGCTGCCCC TCAATGACAG CTCCTGGAGC AGGCTGAGAG AAGACAGATT TGGGAAGATG 240
CGAGGAGAGC ACTTGCAGCT GTTGCCTTAC CTGATGAGCG CCAACCCCGT GAACTATGAC 300
CGGCTTCACG GACTTTCCTG CTGAGACCG TTGCTGCACA CTTTCTGCAT CGTACGCTTT 360
CGAAGCTTTC CTGTCAATTT GCTGGCGAAG TTATTAATGG GCAAGGAGCTT CTTCGACCTG 420
AAGCGCCACG TCCCTGACAA GTTACGCTCC TCGGCGAGCC CGAGGAGAGT GCTGCAAGCG 480
GACGAGAGAT TCTTGGCCAA TCCCAAGAG AGCCCCGAG AGAGAGAAAT GAACTCTTTC 540
GATGAGATT CAGGAGAGAA GTTTGAAAC CCGACAGAGC CTGAGCCGAG GACCCGAGTG 600
CCCTGGACAG CTGATGACAG TGAATGCTCT GAGGACCCAG GGCCTTACGC CAGAGCGGAA 660
GGAGCCAGCA GCACTGCTCG TGAAGAGAG CAGACCGAG GACGCGAGCC TGAAGCCAGG 720
GCCCCGAGCTG AGCTTTGAAA AGGATTCAGG AAGCGGCGAG GAGACTGAG GTTCAAGACA 780
CAATATTTT TGAAGCTGAG TGAAGGAAA ATCTAGAAC ATGAGGAGCA TTAATGAGCC 840
TGGCAGCCTT GGCCTTTCG GGCCTGAGCC AGGACTGAGC TGTCCGAGTT CTCCCAGAC 900
TTTCAGACAG GCTGAGCTCT GTGCTCTGCC TGGGCGCTCT GCGAATGAA GCTTCAGGCT 960
AAGAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA GGGGGGGGGC 1020

(2) INFORMATION FOR SEQ ID NO: 82:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 770 base pairs

60

232

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(*1) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TGAGCCAGCG GGTGGAGGCC GGCCTTACCC GTCTTGGATC TCCCGAGTCC CGCTGCTGCC 60
GCCCCCGCCT CGGCTGCTG AGCCAGAGAC GAGCTGACCG CCAATGCGAG CATTCAAGCT 120
TTGATTAATT TGTCTCTTGG AGAGAGATC GAGCTGATGT TTTTATATGT TGAATGTGCC 180
CTTCGAAAT ACGAAGAAA CTGGCCCCCT TTGTGTCTAT TTTTATACAT CCTTTCAGCT 240
ATTCCATTAAT CCAATACAG AGATTAATG ATGATTAACG ATGCTATGAG TAAAGCTTGT 300
AAGGAACTTC CAACTTTCT TACACAGGAC ATTGTGCTGT CAGCTTTTGG ACTGCCATTT 360
GATTTGCCA GAGCAGACT GATTGAGTGG GAGACTTTC CACTTGTCT CACAGGAAAC 420
ACAATCAAT TTGCAGACTT ACTAGCTTT TCTTGTGCT TTGAGACAA TGAAGACTTC 480
ACCTGCGAC ACTGTGAAA AAGAAATTAAT GAACTATGT CAAATGAACT TCTGTGTATT 540
TGTGTGCAT TCAAGACAG AGAGATGAG GCAATTAATG CTGAATGATA TACGAAAGCT 600
CTTGGGGGTA TTTTATGTGC TCCCTTCTCA CTTTATATGT AAGCATACTA TTTTCAGAA 660
GACTCTGTA AGGATTAATA GAAATTTCTC TTTTGGAAAA AAAAAAAA AAAAACTGAA 720
GGGGGGGGCC GTTCCCATTC TCCCTAATG AATTCATTT TTAAGATCCC 770

(2) INFORMATION FOR SEQ ID NO: 83:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(*1) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

GAAATCGAG CAGCAATAT GTTAAACACT AAGATTAAT GGCCTTCTTA TCCAAAAATG 60
AAGACTGAA TCAATTTCT TCCCTTTCCT TTACAAAT TTACAAATTC AGGTGAGCTCT 120
TTCCGATAC GATGAGCTGA TTGATGAA TGCACACAG TTGTGATCT CCCCACCCC 180
ACAGATTTTC TGGCTTCAT TCGATTGAC CCAAGCCGAG CAGAGGCTGA CTGGGACAAA 240
ACGAACTAT AGGCGGTGAA CCAATCTCT CTCCGTGCCC GGAAGCAAC CCGGGGGGCT 300
TTCACTTCC CAGGAGTCC AAGGGGGGGC GGGTACCCA ATTCCGCCCC TATATGAAAT 360
CGTAAATAC AATTCACAT GGGCGGTCCN TTTTACAAA CTTTCCGTTG AACGTGGAAA 420
AACCCCTGG CGGTTAACC CAATTAAT CGGCTTTCG AAGCAGATCC CCCCCTTTT 480

233

234

c

481

360

5

5

420

(2) INFORMATION FOR SEQ ID NO: 84:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 644 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10

600

15

15

720

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

GCTGGGATAG AGCATGAAAG GAGAACTGCT CCCCTTTCTG TTCTCAGAG TTGCTTATG

60

780

GCTTTATAAA CTCTATATTT GTGAAGGCC CAGATACCCA AATGTCATTT GCAAAACTTA

120

840

TTTTTTTTT TGGACAGATC AGATTCTTAG AGAGAGCAGA TTCTTAGAGA CATTAGCAAT

180

900

CATAGTAGT GAAATTTGTC TAATTTTTT ATTCCATGCT ATTACTGGGC ACTAGTCTTA

240

960

ATTTTTTTT ACMAAATA GATCTATTTT CCTATATAT TGAATTAGAA TCTTAAGTTA

300

1020

GAATTTTATA GAAGAAATGT CTGACAGTT CTATGTATG AGGAGCAAT CAGCTTTTCA

360

1080

GCACCACTT TATCTTTTC CACTAGAGG AGATCTGTG TTGCTTTCTC CTTTGGAGAA

420

1140

TAGCTGCTT GCTTTTATTT TTAATTTCTA AGTTTGGAA AGAATTAAT CTCAAATTC

480

1200

CTTTAGTGT ATTAAATATT TCTATTAAT AGTCAAGGT AAGTTAATTA AGCTTGTTTA

540

1260

ATGATGCCA TCTATGCTT TTCTGTATC TTCAATTTT ATTAATGTG AGTTAGATAC

600

1320

TAAGTGAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAA

644

1351

40

40

(2) INFORMATION FOR SEQ ID NO: 85:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1351 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2527 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

45

45

50

50

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

GGCAGAGTG CCAGAGCGTG GGGCTCTCTG CTCTCTAGT GGGCGCGGT CCGGCTGCT

60

60

GGCTCTGTG CAGCGCGGC GCGGAGCTC CCGGCTATG AGCGGCTTCA GCACCGAGA

120

120

GGCGCGCGG CMTTCTGCC TGGAGTACG AGTCTTCTC AAAAAAGCA AAGGACATA

180

180

TATATCTCA TTTCATGTA TTCCATTTA TGCAGATAG CATGTGTTT ACATGTAGT

240

240

TCAAGTACA CCGTGTCTA ATCCAAAT GAGATTTCT ACAGAGACC CTTTAAACC

300

300

ATGCTACACT CTGAGTGGTG ACGATTATCG TCGAAGCTGC AATTAATAATC TTGCGCTTGG 360
TCTCGAAGAC AGAGAGAGAG AATGTCAGAG CATTCGAGAA GTTCTTAAC TCCTTTCGAGA 420
TCTCGCTTGA GTTGCATCA CAACTTCAAT ATCAAGCTTAC ACTCGAGATC GCAACAATGC 480
AAGCGAATTT CATGAAAGTG CAACTTACG GTTATTGAA GTATTGTGTG ATGCTCTCT 540
CGATTGTGT GAGCAGAGAG ATGTCAGAGG ACTTCAGAAA AAGCCCCGG CAGAGAAAT 600
TAAAGGTTC ACTGAGATCG ATTCTGATA TGAAGAGCA GAGGCCCTCG AGTGTGTCT 660
GAAAACAGAC TCTGTGATG TAAATGACTG TGTCCAGCAA GTTGTGAGAC TTCTACAGGA 720
AAGCGAATTT GTACCTGTG ATGCAATCTTA TGAATGATA GAACTATATG TGCAGAAAA 780
TAACTTCAAT TTGCGAATAA CAGATGGGGA AACATTACCA GCACTGAAAA TTAAATTAAT 840
GGAATATCAG TGGGTGAGAG TTATTGAGAA AGGTTCGGCA ACCCGATTGA ATGCGCTTAAT 900
GAGAGAGAG GATGATCTGC AGTGCCTTCA TTATTGATGT CTTCGTGATG GAGGTGTCAAT 960
TAACTGTCA GTACCTATGAG TTCTGACTGC GACTCATGAA GATTAAGAAA GCGTCAGAGG 1020
CTGTACAGCA TTGTCTGTGA TGTATGAGGG CCGCCGTGTG GCGATTCTC GCAATTCGAA 1080
GTTTTTGGAG CAGAGAGAG AGGAGCGCTG TGCAGAGAG TGGGGAGACA CATGCAAGAA 1140
CGACCCCTAT ATTGAATGCG TGAATGACA AGGAAATGCG CTGATTGAGG GAAATTTTCA 1200
AATCTTGAT CCAATTATTT GCAATGATG TCTTGAATCG TATCTGTCTA CTCTACTGA 1260
GCTTAACAGC AATTTTAAAG ATATGATATC TGAATGATCT TTTCGATTTC AACATACGAA 1320
CGCAGTGCAG AATGAGATG CCGCTTTAAT GCAAGATAC CATTAAGCAGC TTCTACAGAG 1380
GCGCTAACCG CGCCCTGTTC TCTTCTCTCA CCGTCGAGT GCGTCGAGAA AGATATGACA 1440
TGTCTCTTGG ATGTGAGCTA TGAAGAGACA TGCATGAGTG TTGAGAGAG GATTTCTGAA 1500
TCTCGAGAGC ACGATGATG CCAATCTCC ATCTCCATG ATGTATCTG GACCAACTGA 1560
GATTCAGTGG CATTCAGAGC CACGATGAT TGCAGAGACC AACTTTACA TTGTGAGAG 1620
AAGACCTGCT GCGATGCTTC ATTCAGAAAC AGGAAAGAT CTTTATGAGC GAAATCATGCG 1680
TGCAGAAAG CTGAGATGCG CCGCTGTATT AATCACTTGG GAAATGATTC CCGTTGAGT 1740
TGCAGCTTAC AACAGAGAAA AGAAGGATAT GAACTACTAT GACTGTGAGC AGCATGAGAA 1800
CTTTGATATT ATTTTCAAGAA CAGGATGCG GAAATTTGCT CAGAGAGGCC AGAAGACAGC 1860
TGAAGGTTC ATGCTCTCCA AAGCTTGCAG CTGTCTGACA GAAATCTACA AATCTTGA 1920
GAAAGCTTAC GCTGTATACC CAGTCACTGC ACCTTTGACA CATTAATAT AACAGAGGG 1980
GACCAATGAG TGTCTGTGCG CATTTCTTGG TGTGTGTCT TCGAGATGC TTCTAATAA 2040
GAGAGCAATTT TCTTAACTT GCAATGATTT TGTGTCTCT TATGATCT GTTTGAGAA 2100

AGTGAAC ACAGATGTT TTAATGTATC TTATCGACTT AATTAATGTA TATCTCTACA 2160
ATCAATTTT AAAATGTCT TTATATATTA TATTAATCT TGTGTGTAT GATTTTTCGA 2220
ACTGTATTA TTAATGTATA CCAATGATAT TCAATTTAAA TCTGTCTTTT TTCCCTCTTA 2280
AAAAAGAAA AAAATTTACA AACATTAAC TTGCTTAGAC CTGTATTGA GATTTTACA 2340
AAGCTTTGT AACATTTAGA TTTTATTCT ACATTTGAAA TGAATCTGC TTCTTTCTT 2400
CTTTCCAGTC ACTATATGAT CTTCAGACT GTTAAATCT AAATATTTCT TATGATCTGT 2460
GTAGCTCTG AATGAACTTC TTATCTCAT TAAATTAAT TTGTGCTTTC TTAAAAAAA 2520
AAAAAAA 2527

20 (2) INFORMATION FOR SEQ ID NO: 87:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2566 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

25 (4) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

30 CCGAGAAAT CGGACAGAG GAGCGAAAT TGGGATTTCT GAAATCTGTA GCGCCGAGAC 60
CGATCACTT GCGGAGAGAA GATTCAGAC CTACATTTCC CTGCGCTGCT GAAACAGAC 120
CATCTTTCG CAGTGTAAAC CAGAGCATG ACTTAAGCC ACTAGGCCCA AATCTGAGCC 180
TACTCTTCA ACTCTAGAAA ATGAGAGAAA GCAAGCTTT CCGAAATTTG CTGGGGTTTA 240
AGGAAATTT ATGTGAGAT CAGAGATCT TGAAGCAGG CCGCTCTTCC CGAAGCCGCG 300
CTTTGCGAG AAGCGGCCCC TAAATACGTA GAACTCCAT GAAAGCAGAA GCGCCATGAA 360
GAAATGTCT TCAATCAAGG GATCCCGAG TCCCTTGGGA GTCAAGTCCA AAGCGGCGCC 420
TTTAAGACA GCAAGAGAG ACTCGAATA TAAAGATCT GCAAGGAGAA TTTCAGATTT 480
GCGCTTCTCT GAGATGATTT TAAAGCTGC TCGAGAGAG GAGAGCCGCG GTCTCTCAA 540
AAATGTGAA GAAAAAGAG AAGATGAGAA GATTAATCT GCTTAAGACA CCGTCAGAG 600
CAAAATTAAT CAGGAGAGAT TGGCTTCAG GATCTCTCT GCGAGTTTCC CTAAAGCCCC 660
TTTAACTG ACATGAGGG GCGCATGGGG CCAATCTGAG GAAAAAGAA AAGAGAGCAA 720
GAAATCAACC ACCCGAAGC AAGAGCAAT GCGTCCCTTG TTAACTGTG GTCAAGCTCC 780
ACCAAGCCC AAGAGAGAC GAAATGTGA CCGAGAGAA TTCCAGAAA CCGTCTCTG 840
AAGAGTACT ACGAAGGCC AAGCTGTCTA CTCAAGACT TCCCTGAGC GACTTCAGC 900
ATCCATGCG GCGAGCAGC GACATTTCC AAGATTTAC CAGTCAAGC GACAGATCCC 960

237

5 AGCCTACTT CCGAGAAACA TTAACTCTCC GTTTGACCTA AANAAGCCCTG TCAATCAAGA 1020
CAATCAAGAT GGTGTGACCC ACTCTGATGG TGTCTGAAAT CTAGATGAGG AACAGACAG 1080
TGAGCGGAA ACATATGAG ACATAGAGCC ATCCAAAGAA AGAGAGAGA AAGCGGAAA 1140
CGAGCAAAAG AAGAGGTTAG AGCTGGAGAA AAGAGACAG AAGAGAAAG AANAAGAAAG 1200
10 ACHAGAAATA ACHAGAAAT TTAACTAAC AGGCTCTATT CAGTCTACC ATCTTCGAAA 1260
AGCTTTTGT GATGTCAGAG CAGCAAGAA TGAAGTAGC TTCAAGCAG GAGAGCAAT 1320
TGAAATCAT CCGATCAGAG ACACCCAGA AGGAAATAG TTGGCCAGA CAGCAAGGG 1380
15 TTCAATATGC TATATTAATA CAATCTCTGT AAGATTTGAC TATGATTTCT TGAATCTGAA 1440
AAGAGCTCT CTGTGTGCC CTTGAGACC TTATGAGAT GACCAAGAG TATATGTA 1500
20 TGTTCAGAG CAGATGATA TTAGAGCCA CAGTCAGAT GGAAGTGGAG GATATTTCC 1560
TCCACACCA GATGATGACA TTATGATAG GATTGAGAG GAGATGCTG ATGATGCTC 1620
1800
25 CAGACTAGG GTTCAGAGA AAGATATAC GTGTCTCTGG GGAATTTGA AGATGTTAAA 1680
GGGAAAGAT GACGAGAGA AAGATATAG AGAGAACTT AAGTCTCTG ACTCAGACAA 1740
TAAAGAGT TCATCTTTCC GTGCTCTCC TAAACAATG GACATGGAG ATGAAGTTTA 1800
30 GATGATGAG CATACCTCTG ATTTCCCTGT TTCAACAGA GAGATGATC AAGGACTTAA 1860
TGTTCGAAA CTTAGACAG AAGAAAGGA CTTTAGAG CTAANAAGC AGRANAARA 1920
35 ARAAAGAG TTCAAGAAA AATTAAATA TGAATGAGAA ATTAGAGTCC TATATTCAC 1980
TAAAGTTACA ACTTCATAA CTTCTAAAA GTGGGAAAC AGAGATCTAC AGGTAAACC 2040
TGTGATCT CTAGAGTTA TCAAAACC ACATGACACA AAGTTCTCT CAGAAATCA 2100
40 AGAAGGAAA TATGTTATG TGTTCGGAG TTACTTAGG GACATGATG CAGAGATCTA 2160
TGATGATAT CTGATGGCT GATCTATGA CAATGACTAG CACTCACTT TGTCTATCT 2220
45 CGTGTGTTC TTAGTGCCA ATGTGAGTC TGAATTTTA TTGCAATGTT ATTCGATATC 2280
AAGAAATTA ATCCAAAA CCAATTATA TCAATTTGTTA TGAATGCCA ATTATCTTTA 2340
CAAGTGTTT AAGTTTGA CAATGAAAT AATCTCTCTG CTTAATGTT ATCTCAGAG 2400
50 ACTACATAG TGAATGTA GAATATATA ATATTCAT TTCCCTTTGG CTACATATAT 2460
GAGAGGTTG AAGTACTTC TTTTAGCCA CCAATTAATA ATCTCTCTC AAAAAATTA 2520
55 AATAAAAA AAAAAAAA ACTCAGGGG GGGCCCGGTA CCAAT 2566

(2) INFORMATION FOR SEQ ID NO: 88:

60

238

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 540 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

5 GANTTCGCA CGAGCCTTTC TGTGTCTCT GTGGCTCTT TAGTGTGCA CAGGSGCAG 60
10 ACTTGGTGG GTTCAGCAG AGATGGCATG GGCCTCAAGG TCCAGATGT TTACTCTCTT 120
180 GCGGTCTTC TGTATCTCT GGTCTTTGTG GTTCCACAG TTTCTTTGGA TCCAGAGTT
240 AAGGCGATG CTGAGGATG ATGGCTCAT CTCCAGTT CTCTGATG CTGATTTCA
300 GAGTGTCAA AGAGGGTGG CAGACATGT GTGATGCA TTCAGACCC AGATGTGGT
360 GCAAGAGC AGGATGCA CAGCAGGTA GAGCTGTT TCCAGGCCA AGCAGCTTC
420 AGGAGTGT GCGCTCTTT CTGATGTTT TTGGAGTAA GATATGTA GACATGGGG
480 GTCAATGAG TCAATAAAA CTTCAAGAA ACCTCCATG GCATGTTGG CCGAGTGC
540 TATGCTGT AACCCAGCA CTGTGAATG CCAAGTGA AGGATGCTT GAGGCCAAGA

(2) INFORMATION FOR SEQ ID NO: 89:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1863 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

40 TCGACCAAG CTTCCGCGA GATCCCTACC GCAATGACC CCTCTGCGC GCGGAGCTT 60
120 CCGAAGCTC TTGAGCGCC CCGAGCGCT CCGAGAGCC GCGGTGAG GTCGATG
180 CAGCGGAG CCGGAGCCC GCGCCGAG CCGCCGCG CCGTTGAG GCGGCCGAG
240 CCGCGCAT GTGAGTGA GTTCACTC GGTCTGACC CAGAGGAG CCAAGAGCA
300 CAGGCCAG AGCGCGAG AGGCTCAT CATCCCGCC CAGCGCGTG CCGTGAATG
360 CAGGACCA GATGATGG TACAGTTGG CCAAGAGA GCTGTGTT GTTGCATGT
420 CTTTGACTA GATTTATG TTGAGGTGT TATCTAGA GAGCATCT TGTACAAATA
480 TTTTGACTT CAACAGAT ACCTGACTA CTGTGAATA AAGTACATCA AAGATGAT
540 CATCTAAT GAGCCCTCT CAGATGCCC AGCTGCTTC TACCAGACA TTGAGAAAA
600 TATTAATTC TTGAGAGAG AAGAGTTGA ATTATCAT GTGCTGTCC CAGAGTTTC
660 AAGATGAT CTTGCCACA TTGTTATGA CTTTAACAG AAATTCACG CCTATTAGA

	TCCTAACCTG GATTAAGTCT ATGTGATCCC TGTGAAACAT TGTCAATTGA TGTCCACGAG	720
	AAACGACGAG GAGTTACTTA TTACATCAAA GGTGTGAAAC TATTTCCTTC AGTCTCATCT	780
5	GATTCAAGAG CAGATGTGTA TTACTGATCG CATTGAAAC AGTGATCAAC TGGCTTTCTT	840
	TATTATGCA CTGTGTCAAG ACAGAGAAC TTGAAACTCG CAGCCGAG AGACTATTGA	900
10	AGGTATGAG AAAGTGAAG CGAGAAATCG TTTCGCAAT CCGCAATTG AGAACAAATT	960
	TGCGTGAAA ACTTAAATT TTCTTGAAAC AGTGTGAAA AACATTATG AGGAAAATTGA	1020
	ATATCAAGC ATAAACCCAC CCTTTACATT TTGTGTCAATG ATTATTTTTT AAAGTCTTCT	1080
15	TTGTATGAG TACGAAACAG GCGTTTACTA TCTTTCAATC TCTATTATTC AATTAAACCC	1140
	ATTACCTTAA AATTTTTTTC TTTCGAAATG TGTGTCTTTT TATATTGAA TTAGTAACTG	1200
20	TATGAAATGCA TGAATTAATG TACATGTAC CTTAAGTACT AGGAAAGATT ACAATTTCTT	1260
	TAAATCATTT ACTGTGATTT TTATGTTTTA TTACGATTTT CAGGAAAGAG GATTATCTAG	1320
25	AGATTAATCA TTATTAATGCA TACGTAAAA TGAACACAG TACATTAATT GTATGTGTTA	1380
	GTGTCCCTGC TACGTAGTTT GTTATGTCAAT TTGAGACAC ATTTHAATTT TCGTCTAATTT	1440
	AAATGTGCA GTATTTTCAG TGTCAATTAAT ATTAACTAAT TTAGGAAATG ATTTCACACT	1500
30	TTATGTTTTA AATGCTGAG CACTGTCTGT AATTAATTTT TGAAGAAATG TTGCAATTGA	1560
	AGAAATTAAT TGTATTACTA ATTGTATGTA CCGATATCTG TGCATATGAA TGTAAATATC	1620
	ACAAATGTGT TTAACTAGAC TGTGTGTGT TTGTCCGTTA TAAATTAACCC AAAGAAATAGT	1680
35	TTGTGTCTGC AAATCTTAAG AGATTCACA TAAAGAGAG AACATTTTTT TAAAAATCA	1740
	CTTCTAATTA TGAATGAGT AAATTCACAG ATTTTTCTTT TAAATTAATA TAAATGTATTT	1800
40	TAAATTAATTA ACCGATTTCT TTGTGATTAAT ATTAAATGTA CATTTAACCC CAAAAATTA	1860
	AAA	1863

45
(2) INFORMATION FOR SEQ ID NO: 90:

50
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2478 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

	GGAGACGCG CAGAGGTGA GCTGAGCCG TGGTGAAGG GCGGACGAG CAGTCTGTGC	60
60	TGTGGGGGCT ACAGAGAAAG ATCTAATTAT CAGTCACTCG CGACAGTTTC TTATGTGCTT	120

	GTCCCTGTGC AACGCTTTTG CCTTGAGCAA ACCGACGAA AAGAGAGCC GTGTATGTA	180
	TGAGCTGAG CTCATGTACA AGGTTCACA TGAATCTAG AGTTTGAATT ATGACATATTA	240
5	TGCTCTTGG GATCTGAGAG AAGGAAACAG CTTTATCTAG CTGACACAG AAGAGACAAA	300
	GGAAAGCTTT GGAAAGATTT TAAATTAATT AATATGAGAC AAGAGAGGCT TTGTCACTGT	360
10	GAATGAGCTC AAAGATGGA TTAAATTTGC ACAAAAGCC TGAATTTTAC AGGATGTGAA	420
	GGAGAGTGG AAGGGGCAAT ACCTCAATGA GAGAGGCTTC GTTTCTGTGG AGGATGTGAA	480
	AAATGTCAAC TACGGCTAG TTATGATGTA TCGAATGCTT GATGATGAT TTAACTATTA	540
15	ACGATGATG GTTAAAGATG ACCGAAAGTT TAAATATGCA GACAAAGATG GAGACTCAT	600
	TGCGACAGAG GAGAGATTCA CAGCTTTGCT GCACTCTGAG GAGTAAAGAT ACATGAAAGA	660
20	TATATGATGA CAGGAAACAA TGAAGATTAAT AGATTAAGAT GCTGAATGTT TCAATGATCT	720
	AGAAAGTAT ATTGTGTACA TGTAAAGCCA TGAATGAAAT ACTGAATGAC CAGAAATGCT	780
25	CAAGGAAAG AGCAAAAGAT GGAATCTTCC CTGACATAT GATCAATCAG AAGCAAAAC	840
	CAGGACACTG GTCTATGAAAT CAGACCAAAA CAGAGATGAC AAGCTTACCA AGGAGAGAT	900
30	CCTTGCAAG TATATCTATT TTGTTTGCGAG CAGAGCCACA GATTTTGGGG AGGCTTATGT	960
	AGCGCATGAT GATTTTGTAG CTGCGAGAGA ACCCTCATTT CTTCAAAAT AATTATTTTT	1020
	TACAGCTTCT GATTTTCAAT GAAATGTGTT GCGTACTGA GACTGTACT ACAAACTTTT	1080
35	TACAGCATTA AAAGGCTTAA TGAAAACAT CCGTCTCCCA TTCTCTCTCC TCTGTAGAGG	1140
	ACTGAGAGGA AGCCTGTCTT CTGAGAGACA ACTTAATTA GTACACTTGT GTTGTATGAT	1200
40	TTACACTTGG TATTATGAT TAACTATGCG TGTTHAATTT TGTATTTTTC TGTGTGTGCG	1260
	AGTATGATAT GAAGATGACA GATCTGAC TCAGACATGT AGAGAAACAT TACGCTTTTA	1320
	CTCTTGTCA ACCCTTTTGA TGAATTTAAT AATTTCTACT TAACTAATTT TGTAAAGCTG	1380
45	AGATCAATTA GAATATTTCA GAGAGAGAGA AAGAAAAAAT AATATATCTC CAGATTTAAT	1440
	ATTAAAGAG AGAAACCTTA GTCTTGCTTC TCAAAAATG CAGATTTTCA TGAATGATG	1500
50	GGGCGAGTA TTACATTTAG TTGTATGAG TCCAGCACT GAGCTGTCCA TTACTGTGCGC	1560
	AAGGAAATAT CCGTTGTGTC TGAAGAACT TGGGTCGAAT TGAATTTCTT CTTTTTCCCC	1620
	CTTAAAGACT GACTGTGAC TAAATTTGTC AAGCAAGCT GTGTGTGAAA GAGTTAGGCGC	1680
55	CAGTGTCTG AAAATCAATC AAGTATGAAA TGTATATCTT TTGAGAGACT ATGATATGAA	1740
	ACAGGTGAAA AACTAAGGAA AAATTAACAG TGTTTTGGGG GGAATCAATTT TTTTCTGCGG	1800
60	TGTGATATG TTAAATGCT CAGACTTAA TTATTTGCTT TTGAAATGCA CTGTAATATC	1860
		1920

241

242

CCCCATCGG TTCTCTTCTT TCCAGGTGT GCCAAGGAT TAATCTTGST TTCACATCAA 1980
TTAAMATCA CTCCTTTCCA ATCATGTCTAT TGAAGTGCC TTATAGGAAA GAATGTGTCA 2040
5 CTGANTGGGA ATTCTCTTAA GAATCCCTGA GATTAAABAAA AGACTATTTG GATTAATTAT 2100
AGGAAGGCT AGAAGCTCC AGTAGAGTGG GGAATTTTTT CTCTCTCCCT TTCTCTTTTG 2160
GACAAATGTT AATATACAG TATATGTTAT GAGTTTGSTT GCAGTGTCTT TATCTTGTGG 2220
GCTGATTTCC AAAAACACA TCGTCTGAA TTATCCAGGG ATCTCTATAC CTCACANTCC 2280
AATCCATTA CTACAGGCC TTTTCTGTG TCCACTGTGG AGCTTGAGCT CACACTCAAA 2340
15 GATCAGAGGA CTTACAGAGA GGCCTCTTTG GTTTGAGGAC CATGCTTAC CTTTCTGTCC 2400
TTTGACCAT CACAGCCCAT TTCTCTCTCT TTCCCTCTCC CCGCTGCGAA TTCTGTGAGC 2460
CCGGGGGAC CACTAGTT 2478

(2) INFORMATION FOR SEQ ID NO: 91:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2058 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

55 TOGCGCTTCC TTTTGTGTC TTCTCTGTG GCGAGAGGTT TTTCATCACC AAGCTCTCTG 60
35 ATGCGAGTNC CTTACCCGAT ATGTTCAAGA TACTGAGGTA TTCTGTCTGT TCCGAGAGC 120
GAACTGAGAA GCGGACAGAT AATGTGAGG GCATTTGAGT NTTTTAGCAA TCTTCTAACC 180
40 AAGCTCTGTT TGAATCATGT AAGATGTCTC ATGCTGTGGC ATTATAGAAA CAGAAAGTGG 240
AAGATGTGAA AGCTCTGCTC AAGATTTGCC CTGTTTCTTT GCGTTTGATA CTTTACTGGA 300
CAGTGTATTT CCAANTGCAG ACACATATAT TTTTACAGAG TCTTCAATTT AGGATTCGAG 360
45 AATTTTCAA TATATACAC ACTCTTACA CCGTCCCTCC AGCTGTGCTG ACCATGTTTG 420
ATGCTGTGCT CATCTCTCTG CTCATCCCTC TGAAGGACAA ACTGTGTGAT CCCATTTTGA 480
50 GAAGACATGG CCGTCTCCCA TCTCTCCCTGA AGAGGATCCG CCGTGGCATG TTCTTTGTCA 540
TGTGCTGTGC CTTTGTCTCA GGAATTTTGG AGAGTAAGAG GCTGAACCTT GTTAAGAGAA 600
AATCCATTHA TCGAGCCTNC GCGACGTGTG TCTACATATC TCCGATCTGT TCCGTGTGCT 660
55 GCGAGTGGC GCACTACTTG CTGATTTGGA TCAGCGAGAT CTTTGCAGT ATCGAGGCC 720
TGAATTTGC ATACTAGCT GCGCCCAAGT CCATCCAGAG TCCCATATG GCGTGTCTCT 780
60 TTTTCTCTTC TGGGTGTGGG TCGTTGTGTG GTTCTGTGCT CGTGTGACT GTGTCTATCA 840

AMGCCATCG ATGAGTAGC AGTCACAGAG ACTTGTGTA TATTACGGC TGCTATTGA 900
ACTATTACTT TTTCTCTCTG GCTCTATTC AAGGAGTAC CCTCTGTGTT TTCTCATTA 960
5 TTTCTGTGAA ATATGACCAT CATCGAGACC ATCAGCGATC AAGAGCCATT GGGGTGCCCA 1020
CCAGGAGAG GCGCTGACT TCCTGAGGCC ATGTGGGTTT TCTGAGGCTG ACATGTCACT 1080
10 AACTGACTGG GGTCCACTGA GAACAGCCAA GACTTTAAT TCCCATAAAA TGTCTGACTT 1140
CACTGAAACT TCGATGTGTC CTGATTTGAT TTCTCTTTTC CCTCTATCCA AAGGAGCTTG 1200
GTAAATGCTT TACTGTACGG TGTCTCTGG CAGCTGTGGC CCTCGGGGAG GAGAGCTGCA 1260
15 GATTTGAGT ATGTGCTTGG TCATTCAGG TCTGTGTGAA TCTCTAGCT GGGTTGCCCT 1320
TTTTACAGAA ACTCACAAAT CGAGATGTCA AAGTCTTGGG GAATCCACG TGTTAGTTGG 1380
CATCCGAGTT TCTTAACAA ATAGTATCAC CTGCTTCCCA TAGCCTATTC TCACGTGTA 1440
1500 AAAAAAATTT AATTAAGCTT TACTTATATT TAAGAAAGTG AGGATTTTTT TTTTTHAAG 1500
ATAAAGCCT GGTCAAGTGC TCGAAGGATT TTACATAAAT GCCATATTTA TGGTTTCTCT 1560
25 CCTGAGACAA ATCTTGTCTT TCCCATGTTT TTTGATTTAG GCTGTATGTA AACATTTTC 1620
ATCTGTCTCT TCAAAAAGTA CTACTTTTTT AATCCATCAA CATTAATTTTT CTTTCTTAG 1680
1740 GCGAGCATG CATTAAGATC ATTTGAGACC ATGTGTCCCA TCTCAAGCCA CAGAGCAACT 1740
1800 CAGCGGATC TTCAAGCTT ACCTAGTCCG AGTGTCTATA TATAGCTTTA TTTTGTAGCG 1800
1860 ATTCAGACTA AAGACTGATC ATGCTGTGAT GTTAGGAAAA CATTTCTTTTG AACAGAAATA 1860
35 GTGTATTTAA AATTAATTTA AAGTGTAAA TGTGAAGTTG AGCTGTTTGA CCACTGACAT 1920
1980 TTTTGTATTT TTAATGTAGG TGTATCTGGG GCTTCTCCCT TTGTTAATAT TTTTCTGTGA 1980
2040 TTTTGTGCTG TATTTTGGC ATTAATTTAT TATTAAGAGC ATCTCAATAT CGAATTAATA 2040
2058 AAAAAAAAAA AAAAAAAC

(2) INFORMATION FOR SEQ ID NO: 92:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1411 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

GCACAGAGG CGACCCGGGA GAAGAGAGGC CAGAGAGGCG AAGCGAGGGA GTCTCCAGGA 60
GACCCCGGGA CAGCANTGCC CAGCCCCCTG TTTCCAGGCC TTTCAAGTAT ATCATCTCA 120

243

180 CAGAGATCC CCGTAAAGAG AGAATTCAC ATTCCATAGA GATCTCCAT CCGGAGTTT
240 GAGAGTCGA GATTAAATGA ATCTGTTCGC AATACATCA TCCGTATCT AAAAGCTGTT
300 GGGAAAAAT TCAATGANTT TTTTACCA AGGAAAAATG ATACTTTT GAGATATCG
360 GATTGTGGG GCGCTTGAT CCTTGTGTG ACATCCCAT TAAATCCGA AAGAGATCT
420 GAGATATAG AAAAATAG AGGCCCCAA TTTGAGAGG TGTATCAT TGTCTGTTT
480 GATCAGTTA CCAATACCTT CACTAAAA CTCTTGGAG GAAACATC TTTTTCAG
540 AGCTCTGTG TCGTGGTTA CTGTACACT CCGTTAGAG TACAAATCT GATTGCGCG
600 CTGATACCT TGGCTGATC AGAATCTGTA AACTTCATG TTAGCTTTT TGTGTGATT
660 GTATGTGTG CCGATCTAT AGTTGCCGCC ACACTTTCC TTGCTAAG CAGACTCCA
720 AACCCAGAG CCGTACACT TTAACCTGTT TTCTGTGTT ACTTTGAT CAGTGTAGG
780 ATCTCAGCT TTACTCTCA GTAAATCAG AGTGGAAAT TAAAAACAG TGAATTGAA
840 GATCATGA AAGATCAT TCACATAGA CATTGTCTC TGCCTTAT TTCTTAAT
900 TTGAGGTAT TGAATACAG AGTATGAG GAAATTAAG GAGACATA TACACTATC
960 ACCCTTATT TGAAGAACTG ATGTGTAAG GCTGTCTT TTCTCTTA ATGTCAATTC
1020 TTTAAAAAT GATGTGATA CTACACAG TATTAATGC CTCTTAAG GATGATGAG
1080 TCACCTGAT CCAATTGGT GACACAGT GACTTGGAA GACATAGAT ACATTCGA
1140 AGTGAATAG AGTGAATAC TATTTTCAT TTTGAAATA CCAATTCAG TGCAGCTTT
1200 AATCAGTTC CATTAGAT ATTAAATAT GCTCTCTT TCAATATA AATATATG
1260 TCAATACA TTTCTTTTA TTTCTCTCT TTAAGCTTA AAAGCAATG AAGAGCTTA
1320 GAGTGGATT CATAACGA GATAGAAA ACATGATA ACATATTC AATTTTAT
1380 CAGGGAAAT TCAATCTG TTAGAAAA AAAAATAA AACTGAGG GGGGCGGCT
1440 ACCGATCG NGAATATAT CGAAACAT C

45 (2) INFORMATION FOR SEQ ID NO: 93:

50 (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2187 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

60 GCTTTGCTT TTTTGGCG ACTGGGCGC CCTCCGAG CATTTCAC TTTCCAGAG
120 TTTCTGGA CAGCGAGG GGGGTGGGA CTGCATTA TAAATCCG GAGAGGGA

244

180 GCGGCTAAG AGTAAATG TGTGGGCTC GAGAGCGAG GTACATCCC GAGGCTAACC
240 GAGGCCAC CAGGCCAC CTCGTGAG GAAACACT CTGTGGCTG CCGTCTCTCC
300 TGTCTGAG GACTCTGCT CTCGTAGG GGGGCTGGC CCGGAGTGT AAAAATCTG
360 CAGCCAGC AGCTCTGAC ACTGCGAGA AGTTAGCC CAGGCGGCT ACCTTCCCG
420 AGCCAGCG CCGTCCCTT CACTTTAC CAGCCATAG CCAAGACCA GCAATGAGAG
480 AAGATCTG TGTACATCT GATGTGAC TGTCTCTG GCTATCTAT CATTGGCGGC
540 AAGGACCA GAGCTTCCA GGCAGACA GTCTGAGG CCGAGAGCT GCGAGAGAG
600 GAGTCCAG CCGGCTGGG GAGCTGCTG GCTCATCA GCACTCCAC GGCAGGAC
660 GTACCTGGA AGCTGGGAG CCACTTAC GAAACAGCT CAGTATGAT CCGTATGAC
720 TTTGTGCA GAGAGACA GACTACAC TCGAGACT CCAAGATCA CTTCGCGAC
780 AAGCGAGG GCTGAGAT CATACAGAG TGGGCGGCG AATACACA CCGAGATCTG
840 CCGAGATCA CAAAGAGAT GAAATCCAG GAGCGGCGC TGTATGCA CCGCATTTTC
900 TTAGAACAC ACTGGATGA GAATTTTAC CAGAAATAG TGAACAGG TGGCTTATG
960 GTGACTGAT CCAATACCT GATGTGATG ATGATGACC GAGAGGCT CTACATAC
1020 TACAGAGAG AAGAGAAA GCTGAAATC GTGAGATAC CCGTCCCA CAGCTCTCC
1080 AGCTATCA TCTATATCC CCAATAGTG GAGCTCTG AGCGCTTGA AAAGCTCTA
1140 ACAGAAAGC AGCTAAT CTGATGGG AAGATGAGA AAGAGCTCT TGCATCTCC
1200 TTGCGCAG GTGTGTGGA GATGACCAT GACTGAGGA AAGACTTGC TGGGCTGGCC
1260 CTGATGAG CCAATGAGA GAAAGAGCC GACTTTAC GATGTGAG CAGAGAGAC
1320 CTGTACTG CCAAGTGT CCAACACAC GCTTTGAT TGAACAGA TGGCAACCT
1380 TTAGACAGA TTACGGGCG AAGATGACC ACCGAGAT TCAAGCCCA CCAACCTTC
1440 ATTTCTTAT GCGGAGAC CAAAGGCTC CCGTATATC ATTGGGCGC TGGTCCGACC
1500 TAAAGTAC AAGATGAG AGCAATTA GCTCTAGG TGCACAGG ATGCGAGAG
1560 GATTCAGAG GCTCTGAGA CAGATGGTG CTATTTGGCT TGGGGGAGG GTGAGTACC
1620 AGCTTGAAT ACTGATAG GTGGGTGGA AAGCAGAC GGGGTTCGG TGTGCTGAG
1680 CCAATTCCT AGCTAATTT CACTCCACT GATCAATGG CCAAGTACC ATGATCTGA
1740 GCGCGAAG TGCATCTCT GTGGACTTG GACATATTC ATTGTGCT CCGTAAAT
1800 CCAAGTCA GCTTCCCTA ATCAGTATC AATTTAAG CAGATACCT TCTCATCT
1860 GAGACAAAT TGAAGTAG GATGAGCA GCGCTCTCT GACTTAAG CACTGACCT
1920 GCTTCCAG CTCTATCCA ACTCTTCCA ACTATTAAC TAAATGCT AGCGCTGGG

245

ACCAGGAC CCAGAAATCA CTTGGGCGCA GTGAGGCGGA TTGAGAGCA GTTCCGAGCA 1980
GGGGCTTCTG GCGACTCT GTTCAGAG CATCTGTCT GCGTTTGG GATCACTT 2040
TTTCTTTGT TTCTTCTTT TTATCTCTT CAAGATAGG GAGGAGGG GGAACATGAG 2100
CCTTTGTC TATCATCCA AGACTTAT TTGACATTTT TTTTTCAT AAACTTTTC 2160
CATGACAAA AAAAAAAAAA AAAAAA 2187

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 757 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 94:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

25 GACAGTACGG TCGATTTCC GGGTCGACC AGCGTCCGC GAGCGGTGAA GAAGTGAAG 60
ATGCGGTGG CCAGGGCCGG GTCTTTGGA GTCCAGTGG TCGAAGGGC ATCCCGGAC 120
GTGATGCCGC TGGGCGACG GACAGCCTCC CACATGACCA AGGACATGTT CCGCGGGCCC 180
TATCTTAGA CCCCAGAGA AGCGGCCGCC GCCCCAGA AGTATATAT GCTGTGAA 240
GACTAGAAC CTTACCGGA TATGCGCTG GGTATGGG ACTACCGGA CTTCTCTGAC 300
CGCTCAGC ATGAGAGA TCCATGTAT AGCTGGACC AGCGGGCTT GAGTTGAC 360
TGGGTGAAC CGATGCACCT GCACTAGAC ATGTACAACA GGAACGCTGT GATACATCC 420
CCGACACTG TTCTTTGCA TGTATGTGT ATGAGCTCT TCGTTTCTT GCTTTCTATG 480
ATATTCATGT GCTGGGTGG GAGCTGTAC CTTCTCTAC AGCTGTGG ACCAAGCAG 540
TATCTTACA ATATCTGTA CTTGAGACA GCGGTGATC CTTCCNAGA ACCAGAGCG 600
GTGGTTCACT ATGAGATCTG AGGAGGCTTC GTGGCTTTT GGGTCTCTTA ACTAGGACTC 660
CCTCATCTCT AGAATTTTAA CTTTATGAA ATCCCTAATA AAATCTAGTG CTGTGTAAA 720
AAAAAAAAA AAAAAAAAAA AAAAAAGGGG GCGCCNN 757

(2) INFORMATION FOR SEQ ID NO: 95:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2194 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

246

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

5 GGCAGAGCA CTCCTGCACT TCCACACCC CAGACCGAA CTTGGCTTGG CTNAGCCCT 60
CCCACTCC TCGGCGCTGA CTTCTGGTTT CTTGCGGCT CCGTGGGCC GAGCGCGGA 120
CAGAGCCCC TTTCGGCT GAGAGCTCAT CGACTTCC AATCACTTTC CGAGTCTTT 180
CCCTTCCTC CGGCGGTGC TGGTCCGAC GCGGCGCTG GGTCTCGGC GGTATTTGCT 240
GGTAAAGGG CTTCTCTCG GGTGGCCCG GCGCTCTCT CTTGGCTG TCCCTCTTTC 300
CAGAGTCC GGGCTCTCT CCGAGTGA GAATATGGA CTCCCTCCG GAGTGGCCG 360
GAGCAGTCC CTTGCTCTG GAGCGGCGG TGTCTTGA GAACCGGA GCCCGTGGT 420
ACCCCTGGG ACCCGTTTG TTTCGGTCC GTTTCNAAC ACTAAGAAAT CGAAACTCG 480
CGGCTTGG GGGGCGCTA GTAGCCTTG CTTCTGGTG TCAATGATC ACTGTAGA 540
GATGATCT GTATCTGAA TCATGAAAA GCCATAGA GAGATACAT GACTCCAGT 600
TCATATAT CAGGATGA ATCTGTCT TCCATTTTG CTTCTGTAC TCATATGAA 660
GACATAAA AAGCACTTA GATTCAGAG AAGAGAACT CTTGTTAA GAAGAGATA 720
AGATTTTGG AAAAAAGCT AATAGCTGA TTTCAGAG AATCAAGTTC CTTGGAGCA 780
GAGAGTAA ATAGGCTTA TATGATAT CAGAGGTTT GCATGATAG AGATATTTG 840
AAGCAAAAC TGCANAAAT GATTAAGAC AATCTGAAT CTTTGAAGT ATTGAATGAG 900
CAGTCAAT CTAAAGAGT AGACTCTTC CAGTGAAG CAGAGTGA AACTCAGAG 960
GTGATAGA ATTTAATCC ACCTTCATCA AACTGGAG TGAAGATT GAGCTGTGAC 1020
CTGAGATCC ATGTTTGA ACAGAGCTG GACTGATGA GGAAGAAAT TAGCATCTC 1080
AAAAAGAAC TACAGAGC CAACAAAG GATCCATATC AGGAGACAA TCTGAGAGC 1140
AGGATCTCC AAAAACTAG CATTTCAAT GATATATGC AGCATGATA CTGGGACTG 1200
AAGAGGAAA TGTCTAATTT ACATCTGGT ACTCAATAC AAGCTGAAT ACTAAGAAA 1260
CTGAAAACCT CAATGCAT CAGAAAGCC TGTCCCTTG TAGGATGAG TGAAGACTT 1320
GGAAGAGCA GGCANAACT GCATTTGAT AATTTACTG CACATACAC AAGCATCCC 1380
CCTCTCTAC CAATGGCA AGCTCTTGT CATACCAAT CTTCCCTTT ACCAGAGAT 1440
GTAAAGTTT TATCAGAA AGCAATCCT CATTATGGA CAGCAATGA GAGATCAAT 1500
CCTAATGAT GTACATGCT TCGAGAAC AGTTCTTAT GCAATATTC TCTGAGAGC 1560
AATTCCTGG TATTTCCAG TCTCTTAAA TCAGTGA GAAGTTTG GGAAGTAAA 1620
ACTAAACTT TGCCTTACC CACTTCCA CACTGCAT ACTTGATCA ACATATACAG 1680
AACTGCTTT ATAGAAATTA ATTTGGAAG GATTCAGAT TTCACATGA GGACACTTAT 1740

247

1800
1860
1920
1980
2040
2100
2160
2220
2280
2340
2394

CTCTTTCAGT GGTCTCTCCCA AGAATTATTT TAACTAACTG AATGGAGATT TTGATTAAAA
TTTTCAGAGT GTCTTCAAGA TCTATATTGG AACACACTGT ACATATATAC AAAAACAAC
ATACTGGATT TTCTATATAG AAAAGACAAC CTCTAGCTCC ATATCTTAGAG AATCTGAATT
ATGTACTATT ACTATATATT AACTAACTTT AAGCTATTTA AAAAACTCTG CCTCTATAT
TAATTTAAAA ATTATGCTCT TCAAGAGATT GGAATGGAG ATTGTAAAGC TGGTTTTCAC
AAATGTGAAA TGTCTAAATA TCTGTTCATA AAAATTAAGA GAAATATATG TTCTTCAAAT
TGCATATAG AACAAATGCG AATGTAGATA GATTACATTT CTGTATTTAT AATGCTTAAA
GAATATGAAA ATATATATGA ATAAAAAGAT CTATAGTTAT ACATATTTTA TATCTATATG
CCTTTCATAA TTATAGATTT AATATGATTT TTGTGTACCA GTTATTTGTT GATTAATTTT
TTTTHGAAAT GAAATTTGAA TTCTAAAGCT TCAACAAAC TGATCTATGA ACCCAACTTT
GTCCACACAC ATTCTTAAAG TCTATATTAG GAAAAAAA AAAAAAAC TCGA

(2) INFORMATION FOR SEQ ID NO: 96:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 672 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(3) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

60
120
180
240
300
360
420
480
540
600
660
672

AGTCTCTGTG TCCCAAGCT GAGTGCCTT AGTGTAAATG CATCTCACTG CAACCTCCAC
CCCCAGTTC AAGCAATTCG CATGCCAG CGTCCCAAGT ACCTGAATTT ACTGGCATGC
ACACACACAC CAGCTGATG TTATTTTAT TTATTTAATA TTATTTTAT TTAGGTGTTT
TTTTTTTTTT TTTTAAAC GAGTCTTGC TCTATGCC TGGTATGAT TACGTGAT
TACATATCTG GATGATCTAC TGAATGTAC TCACTAGAG TCAATCTTC AATTAACATC
TCAATCTCTG CCTCTTGGCA GATACATCTG GGAATCTTAA GGGGTGAGGG ACTATCTCAC
AAAGAGAT TTACCTCTG TTTTAAATA AACGCAATTT CTTTTCTTA KAAAAATGGG
AAATTTTCA ATTCTTAT ACAGGACAC TGAATTAACA AAGAGAAAG TGTCTGGTTS
GAATTTGGCA KCCCACTCTG GATCTCTCC TACAAAAATG GAAAAAGAAA GAACGTGAT
AAATTAACA AAGCAGAAA AKATTCTCT TTCTAAAG GAAAAAGAT CCCCAATG
CCCAATAACA TCACTGGGG ATTAAGAGAA RAATCTCTCT YCTTGGACC CCCAAACAAA
GCTTAATTAC CC

248

(2) INFORMATION FOR SEQ ID NO: 97:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1419 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(3) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

60
120
180
240
300
360
420
480
540
600
660
720
780
840
900
960
1020
1080
1140
1200
1260
1320
1380
1419

TAAACAGCA ACACAGATA TGAACAGAT GAACTTAAA ACATATGCTG GTGAAGTCCA
CTATATGAGA CAATCTTAT AATTTCGAAA CTGTGTCTTA GTATACAGTG ATCACTGCT
CTCTGTCCA GTCAATCCAC ACATTTCTCT ATTTHAGCT ATTAATATAT AGAAGAGAAA
TGGAAAGCT TACTGGAGC TGAATAATGA ACCTATATAT ATTGCTATAT TTGCTATATC
CACTATTTTC AATTAAGTGT GTACCATATG TACATTTAAA TATATAATAC ATTAAGAGAT
GTACAGAAA TACCTTTAT TGAATATAT TACATTTAT TTATCTCTTA GCATATATAT
ATCTATATG CTCTGTAAAG GCTTTAATA AAAAGAGTCC ATTAATATCTT CTTTATAAAA
ATCTATATG GTTTCATATC TCCCAAGATG TTTTAGATAT AAAATTTTAT GCAAGAGATA
TTTTHGAAA TCTCTCTTGT TCTGATAGAG TTATACAGCA TATTTAATTT TATCTCTGCA
GAAATCCAC AATCTAGCT CTAAACACAC TTGAATATCT ACAGCAATAA TCTGTACACA
TAAATGCCA AATTAAGAG TTGGATGCC AAAACCCAG TCCGTACACA ATGTATTTCC
TCAAAAGCT GCTATACCA TGAATTAACA AAAACATCTG TGTTTTCTTA AACCACTTT
TCTTTTAAA TGTCTCTAT TGTTTGATTT GGTCTGTCT AAAATTTACA ACCTAGCCCA
ATTAAGCTG AATCAAGAC ATTTCATCCA CCAATATCAT GTATAGATAT TATGTATAGA
AAATTAATA AATTAAGCT CTATCTCTG TGTCTCTGT TATCTCTTA TTTTTCGCG
TTATCTATAT GATTTATAT AAGCAATTT ACCTTCGAGA GTTCTCATG CTATCTTTTG
GTCTGATTT TGTCTCTAG ATGAGATAT TTCTTATAG TTCTCTCTT GCAACAGCAAT
GATCTATTT CTATCTATTC TCAATTTGTT ACAAATCTG GATATATCT TACTGTATTA
ATATATCCA AATTATCCA AAACAAATTA TTATATATG CTTTTTCACA GTCTTTGCT
AAACATGTA GTATATGTA ATGTCTCTT GAAATTAAG ATACATCTCT ATTAAGAGACA
GTCTCTTTC ACTCCAGAGA AACTTTTTA AATATGACA CTGAATTTT ATTGCATTT
AGTCACTGA AATGAGATA AAACATACA TGAATCAAG TTTTTHAGG CAGATGATG
TGTCTCTTA CAGATTTAG CAATAGCTT TAAATTTAG TCAATCTATA TTCTATCTGA
TAAATAGCT AACATATATC AATTAATAAA TCAAAAAA

249

250

5 (2) INFORMATION FOR SEQ ID NO: 98:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1487 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

15 GCGACCGCGC CCGTTTCAGC TAGCTGGCTC GCTGCTCTCTG CTTCCTCTCT GCGCGCTGCG 60
CATGGCKMVG GCGTTGCGCG CCGTGGCGCG GCTCGAGCGG GCTTGGCGCG CCGGTACCGAG 120
CAGTTGCGAG ATGAGAGAGA GTCTGAGAA CCTGACACAG CTGCGAGTGA TGCTCTCTCA 180
CGTTACAGCA GCATTCTCTG AGAGAGCGCA GTTTTCCACC TATTTCCTCTG GATATTTTGA 240
TGGTCACTAC TGCTCTCTGT GGGTGTCTCT TGTTTTAGGC TTCTCTCTCT TTCTCAGAGG 300
ATTATTCAMT TATGCGAAGG TTGCGAAGT GCCAGAAACT TTCTCANAATC TCCCGAGGAC 360
CAGAGTCTCT TTTATTTATT AAAGATGTTT TCTGCGAAG GCTTCTCTCT ATTTATGAAT 420
TCTCTCTCAA GAAGCAGAG AACACCTGCA GGAAGTGAAT CAGATGCGAG AACACAGAGG 480
AATATCACC TGCTTTAATA AATATAGTA CTGTGCAAAA GATCATTTCT CTCTATTGT 540
TCTTAGGTGT AATATTTTAA TAGTTAATGC AGAATCTCT ATCATATGAA TCAATAGTGG 600
TTAATGTTTG AAAAACTCT TGCATCAAG TCTGTGATGT ATTAATTAATG CCTATATAT 660
TGTTTGTAGT CATTTTAGT AGCATGAGCC ATGTCTCTCT AGTCGGTAGG GCGCGTCTCT 720
GCTTATTTCA TCTCTCATCT CAAATGAGC TTGCGAATTA ATATGTAAG AGATGATATA 780
TGCTGCGCAT TTTAAGCGG TTTTCTCAAA AGTTAAACTT TTGTTATGAC TGCTGTTTTG 840
CACATAATCC ATATTGCTG TTCAAGTTAA TCTAGAAATT TATTCANATC TGTATGACAA 900
CCTGCGAGCA AATCATAGT GCAAAATATC ATTTAGGTG TGCTCAAAA TAAGTCTTTA 960
ATTGTAAT ATTAAGCATT AATTTTAT AGCCGTATAT CACAATCTG CCGTACCTTA 1020
TTGTACTTAA GCGATTTCTAA AGGTGTGTC ACTGTATATA ACAGAAAGCA CTAGATACAA 1080
AATGAAGCTT AATTACTTAA ATGTATTTCT TGACATCTCT TCTATATTTA GGTCTCTTCA 1140
CCCCACCC CACCCGCC CCGCTATTT TCTTTTGTG TCTGTGTGAT TAGGCCAAG 1200
TCTGGAGTA AGGAGAGAT TAGTACTTAA GGAGCAAGA AAGAAGTAGG TTGGAATTT 1260
TGAGATGATC CCTAACHAC TGTAATCTT GCTTTTCAA TGTGTTAGCA GAACCAAGTG 1320
GGTTATATG TAGATGATG TCTTTTCTG CCAAGTGATA ATTCATCTG GTTTGCTATG 1380

5

(2) INFORMATION FOR SEQ ID NO: 99:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1653 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

15 GCGACCGCGC CTTTCACTA GCTGCTGCG TGGCTCTCTT TCCCTCTCTG CCGCTGCGCA 60
TGGCTTGGG GTTGGCGCG CTGGCGCGG CTGCGAGCGC CTGCGSAGCC GGTACACGCA 120
GTTGCGANT GAAGAAGAT CTGAGAAACC TGAACAGGCT GCACTGTGATG CTCTCTCAC 180
TTACAGAGC ATTCTCTGAG AGAGCGCACA TAAATTTGAC TACAGGATG AGTCTGGGTT 240
TCCAAAGCCC CAACTTTACA ATGTAGTAC AACACTGCCC AGTTATGATG AAGCGGAGAG 300
GACCAAGGCT GAAGCTACTA TCCCTTTGCT TCTGCGGAGA GATCAGGATT TTGTGGGTGG 360
GGATCAATTT GATGATGCTG ACCAGCTGAG GATAGGAAAT GATGGGATTT TCAATGTAAC 420
TTTTTTGAG GCAATCTCTT TTAATCTGAT TGGGTTTTTC CTGTCTTTT GCTCTGACAC 480
TTCTGCTGCA GGAAGGTATG GGGCAATTC AGAATTTGCT CTCTCTCTAA TTAATGAGAT 540
CCTGATGTC AGGTTTTCCA CTAATTTCCC TGCAATTTAG AATTTCTCT CTAGAGACAA 600
GAGACACT CTAGGAGTG AATCAGATG CAGACACAG AGAATATATC ACTGCTTTTA 660
AAAAATATA GTACTGTGA AAGATCATT TCTCTCTAAT TGTCTCTAGG TGTAAATTT 720
TATAGTTTA TCGAGATTC TGTATCATT GATCATTAG TGGTATATGT TTGAAAAAGC 780
TCTTGCATC AAGTCTGTGA TGTATTAATA ATGCTTTATA TATTTGTTGT AGTCAATTTTA 840
AGTACATGA CCAATGTCCC TGTAGTGGCT AGGGGCGAGT CTTCCTTTAT TCACTCTCCA 900
TCTCAAAAG AACTTGGAT TAAATATGTT AGATATGTA TAATGCTGCG CATTTTAAG 960
GGGTTTCTC AAAGTTAAA CTTTGTGTTT GACTGTGTTT TTGCACATTA TCAATATTTG 1020
CTTTCAGAT TAACTAGAA ATTTATCTAA TTCTGTATGA ACACCTGGA GCAAAATCAT 1080
AGTGCMAAA TACATTTAG GTGTGCTCA AATATAGTCT TTAATGTGTA AATATAGC 1140
ATTATTTTT TATAGCTGCT ATTCACAATT CTGGGTACC TTATGTATCC TAAGGATTC 1200
TAAGGTGTT GTACTGTAT AAACAGAAA GCACTAGAT ACAATATGAG CTTAATTTACT 1260
AAATGTAT TCTTGAGACT CTTTCTTAA TTAGCTTCT TCAACCCGAC CCCCACCCC 1320

251

1380 ACCCCCTTAT TTTCCCTTTT GTCTCTGTGT GATTAGGCCA AAGTCTGGGA GTTAGAGAG
1440 GATTAGGTAC TTAGAGACCA AAGAAAGAT AGCTTGAGAC TTTCAGATAG ATCCCTTACA
1500 TACTTACTTA CTTCCTTTTA CATTGTGTGA GCAAGAAACA GTGGGTATTA ATGTAGAAAT
1560 ATGTGCTTTC TCCCAAGATG GTTATTCATC TTGGTTTGCT ATGTAAAC TGTAAATACA
1620 ACGAAGCTT AATTAATATC TCTGTGTAG CACTTTTAA TAAAAAAA TAAAAAAA
1653 AAAAAAAA AAAAAACC GGAGGGGGCC CCN

15 (2) INFORMATION FOR SEQ ID NO: 100:

(1) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1145 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

60 TTTTTTTT TTTTTTTT TTGACTGAC TAAATGCTT TTTTATTGA GAAAGCAGA
120 ATTACAAAAG ACTGCCCTT TCTTGGGATA TGACTGTCT AGCAAGAAC TGAAGTAC
180 TCCAGAACTG ATGTGGCTCA GGCACCTGG TTTTAAATCC TTGAAGATCC GGCATTTGCC
240 TTACGCAAAA GGTACCAATT TGAGTCTCT CTTTACTTAT TATGTCTGC CCAACACTA
300 AATTGTATAT TTGTTTTCT CTAGTTTGA CAGGTCTGA ATTTTTCAT TTATTTCTCT
360 TTTCGCAAC AGACAGACTT GATCTGTGAA AGACAGCAA ATACACTGAC AAGATTTTAC
420 CAGATTTCT AATATGTAA AAGAAAAC CCCAAAGAC TGAAGAAAT TTACACACA
480 AATTTCATT GTTCATTTGA GCACTATTGG TAAATAAATA ACAAATGTTT GTGCATTTT
540 ATGTGAAGAT CTTCTCTGTA TTTCATTTGG AAGATGAC AAGAGTCTG CTTCCTCAT
600 TTTCCTTCCC CTTCCTTTTT TGAAGGAG TTTCGCGAG CTTATGCA GAATATCTGA
660 CTGTTTGA GAAGATATT GGCACATCT CTGAATGTT TTCCAGGTT GTGTATTAC
720 TGAGTTTAT CTTCGAGAA TGAAGAAC ACTGTCACT GTTTGTAG ATTTGTAT
780 AATGTGTAC ATTTTTTTA AATTTTGA CATCAGTA ATAAAGTAT GTATGTACA
840 ATGTATAT ATTTATATA TGACATCAT TTGCAAAAT GTTCCCTG CTGATCTCA
900 TTTTAGAG GTTGCATAG ATCAATATA TGAATATGG AATTCCTGA ACTGTGTG
960 AAGGAGCTTT GTGGCTGTT GCACTAAAG GGCAGATTT TCAAGACCA AGAATCCA
1020 TACCAAGATG AATGTATAG GACTTAAAG AATGAACTG AAGATTTCA CTCTGCTGT

252

1080 TTACACACA GCTTTTGA GAAGAGAA AAGATTTA TCTTATTT TGTACACA
1140 TAAAGCTTC TTCTTTTGT AATTAAGTG AAAAGCTTC CTAAAAAA AAAAAAAA
1145 AAAAA

10 (2) INFORMATION FOR SEQ ID NO: 101:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 734 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

60 TACCGAGCG ATTCAGAGA GTTAAATTTA GTCTTATAT TTCACTTA ATTATTAACA
120 AAGAAACAAA TAAATGAGAG GGCAGCTATT AGCATTCCT TATGCAAAAC ATTGGTTAC
180 TCCCTTTTA TACACTCTTA TCAACAGAC TTCAACATG TATTACATCT CTTCACCAT
240 CCTGTGTGA ACTCAGTTAA AATTAATCT TACTAGAAA TTTTATGCA TTACTGACA
300 AATATTTCT TTTTAAAGTA GTTCTTCCA TCTTATCT GACTACCTTC CAATATGTT
360 TCCCTTTTG AATGAGGTT TTTTGTTTT GTTTTGTCT CTGAAAAAT CATACACTT
420 TGTCTTTCA TTCTTTTCT GTTTTGTAT AAGATGTCC CTTCGCCAA ATGAGAGAG
480 AATGTTTAA TTAATCTTT TAAATTTAAA TAAATTAAT CATTTATAT ATTCAGTGT
540 AACATTTAG TAACTCTTG TAAATTAAG GTTCATAT TTTACTTTGA GATTTTTTC
600 CCTTACTAT TGTGTTTTT GTACTTTAAA ACTTATGGG AATATCACT GTCTGTCAA
660 GAAGACAG TAAATTTAC TGAATTAAT TGAATATCC ACTGACAG GCATTTTCTA
720 TATTAATAAA ATGTGTGTA CTATGTGTA AAAAAAAA AAAAAAAT CGAAGGGGGC
734 CCGTACCT ATTA

45 (2) INFORMATION FOR SEQ ID NO: 102:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 713 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

60 CCGCGGAGAC GCTGTCTG CTGCCAGAC CCGAAGACC TGTCTGTG CCGCGCTTC

253

CTGCCCCGGG CCGAGTCATG ACCCTGGGCC CTTCACTCCT CCGCTTCGAT CTGCTGCTGC 120
TGTGCTGCTCT CAGTCCGGGG GTGTGCGGGG CTGAGGCTGG GCTCGAAGCC GAAGTTCGGG 180
TCTCGACCTT CCAAGTGGAG ACCCTGGTGG AGCCCCCHAGA ACCATGTGTC GAGCCCCGCTG 240
CTTTTGGAGA CAGGCTTCAC ATACACTACA CCGGAGCCTT GGTAGATGGA CATTATTTTG 300
ACACCTCCTT GACCAAGAC CTTCTGGTTA TAGACTTGG CCAAGAGCAG GTGATTTCCAG 360
GTCTGGAGCA GAGTCTTCTC GACATGTGTG TGGGAGAGA GCGAGGGCCA ATCATTTCTT 420
CTCACTTGGC GTATGGAAAA CCGGAGTTTC CACCATCTGT CCGAGCGGAT CGAGTGGTGC 480
AGTATGAGCT GAGGCTGATT GCACTATGCC GAGCCACTTA CTGGCTAAG CTGGTGAAGG 540
GCATTTTGGC TCTGTGAGGG ATGGCCATGG TGCCACCCCTC CTGGGCCCTA TTGGGTATCA 600
CCTATACAGA AAGGCCAATA GACCCAAAGT CTCGAAAAG AGGCTGAGG AAGAGAAAGG 660
AAGCAGACC AAGAGAAAT ATTAATATAT AATTTTAAA AACTTAAAA AAA 713

(2) INFORMATION FOR SEQ ID NO: 103:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

CCGATGTGGA CATTATCTTG TGTATCCCA TGTTCCTGG CTGTACTGCT ATGCCCGAG 60
TCATCTGCTT GCACAGAGC TCTTCAGGA TGCTGTGTTG CCGAGCATGG GGGCCCTCAA 120
CAAGATCAAC TTCAACACC GCTTTGTGAT GAAGAGGCTC ATGACCATCT GCTCTGGCAC 180
TGTGCTGCTC GTTTTCAGCA TCTCTGTGTG GATCATTTCT GCTGTGAGCG TCGGTGTCTG 240
TGAAAGTCTT GAATCAACAG CCGAGCTTC TGGCTATCA CTTCTGTGTT GGTACCATGA 300
CCAGCAGGAC GTAACTAGTA ACTTCTGGG TGCCATGTGG CTATCTCCA TCACATTTCT 360
TTCCATTTGT TATGGGGACA TGGTGGCCCA CACATATGTT GGGAAAGGTG TCTGTCTCCT 420
CACTGGCTAC ATGGGTGAGG GCTGCATCTG CTTGTGTGTG GCGTGTGGG CCGGAAGCT 480
GGAACTCACC AAGCGGAGA AGCAGTTTCA TTAATTCATG ATGGACATCT AGCTACACCA 540
GGGATCAAG AATGTTGAGC CCAATGTCTT TSGGGAACA TGGTTATCT ATTAAGACAC 600
AAGTGTGTTA AAGAAGATTG ACCATGCCAA ATGTAGGAGAC ACAGAGGAA GTTCTTCCAA 660
GTATCCACCA GTTAGAGGC GTCAAGATGG AACAGAGGAA GCTAGTGCAC CAGCCAAACA 720
MTCTGTGGA CTTTTCACAG ATGCAGATG TCMGTATGTA CTTAATACAA GAATCATG 780

254

ACCGAGGCGA AGACCTGGAG AAGCAGATTG CGAGCTGGA GTCGAAGCTG GAGCATCTCA 840
CGCCAGCTT CAACTCCCTG CCGCTGCTCA TGGCCGACAC CTTGGGCCAG CAGCAGCAGC 900
AGCTCTGTGC TGCCATCATC GAGGCCGGGG GTGTCAAGCT GGCAGTGGGC ACCACCCACA 960
CCCAATCTC CGATAGGCCC ATTGGGGTCA GCTCCACCTC CTTCGGGACC CGGTACACAA 1020
GTTCAAGCAG TTGCTAATA ATCTCCCA CTCGAGAGC ATTAAAAAAA AAAAAAAAAA 1080

(2) INFORMATION FOR SEQ ID NO: 104:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 489 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

GGCAGAGAG GCTTTGAGC ATTTTGTCT GTCTCCCTG ATCTTCAGCT CACCACCATG 60
AAGTTCTTAG CAGTCTGCT ACTCTGGGA GTTTCATCT TTCTGTCTC TGCCAGAT 120
CGACACAGC CTGCTCCAGC TGACAGTAT CCAGCTACTG GTCTGTCTCA TGATGAAGCC 180
CCTGATGCTG AAGCCACTGC TGTCTGAACC ACTCGACCA CTGCTGTCTC TACCCTGCA 240
ACACCCCTG CTTCTACCACT TGTCTGTAAA GACATTCAG TTTTACCCA ATGGGTGGG 300
GATCTCCGA ATGGTAGAT GTCTCCCTGA GATGAATCA GCTTGAATCT TCTGCATTTG 360
GTCAACTA TTCAATGCTT CTGTATTTT ATCAACTAC TTACTCTCC TACGATATCC 420
CCTTATCTC TATTCAGTTT ATTTCTTTC AATTAATAA TACTATGAG CAACAAAAA 480
AAAAAAAAA 489

(2) INFORMATION FOR SEQ ID NO: 105:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 640 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

GGCGTCCCG CTGTGTGTT GTTCCCATG GAGCTCCGT AGCGACCCA GCACAGCCAG 60
GAGGTCCGG GATGAGCTCA GCGCGGCGG ACCACTGGGC GTGGTTGCTG GTGCTCAGCT 120
TGTGTGTGG ATCAATGTT CTTAGATTC TCTTCCGCTC CTTCATACC TTCATGTCA 180

255

240 GGGTGTGTCA GAGAGAGGCG GAGCGAGAT CAGAGATGAG AGCGAGATC CAGAGATGAG
300 AGCAGAGAGT GTTCCAGATC AACATGATCG AACAGTTTCC CAGATATGCC AGCGTGGAAA
360 GAAAGATCA CAGATGAGCG GATATGCTCA AAAACCATCT GAAAGCTTGG AACATCTAAT
420 TACCGAGAT AAAATGGGTC ATATGATGTC GTTCTACAT ATTGAGAGCT GCGCTGATGA
480 TGTACTCAT TTGAGAGTAT TATCTGTCC GTGTGGGCTGT GGTGCCAGAT AAAATGATTA
540 CCGTTTACAG GCGCTGGTGG CTTTCTCCAT TAAAGTACGA GGTGGTGTGG GATATACAT
600 TGGATTATAT GTGTACAAAT TGTCTATATG TGTCTACAG TGCATGAGC AGAGAGTGT
640 AACGCGGAG TTAAAGAGG TTTCCTTTCC AGTTTAAAT

20 (2) INFORMATION FOR SEQ ID NO: 106:

25 (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1529 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

30 GCGCAGAGAA TGGAGCTGCC GTAGCGAGAC CAGACAGCC AGAGGCTTCC GCGATGAGCT
60 CAGCGCGAG CAGACACTGCG GCGTGGTTCG TGCTGTCTAG CTTCGTGTTT GGAATCGAAT
120 TCTTGAAGAT CCGTCTCCCG TCGTCTCAT CATTGATATC CAGAGTGTCC CAGAGAGAG
180 TCTTGAAGAT CCGTCTCCCG TCGTCTCAT CATTGATATC CAGAGTGTCC CAGAGAGAG
240 CCGAGAGAG TGCAGATGAA GAGCGAGAT CCGAGAGAGC AACGAGAGCC TCTCCAGAT
300 CAGATGATG GAGAGATTTG CCGATATGTC CAGCGTGGAA AGAAGATCA AGAAGATGAGC
360 GATATAGCTC AAAACCGATG TGAAGCTTCC GAGAGCTGAA TTAGCGAGAA TAAATATGGAT
420 GATATAGCTC GCTTTCCTAG TATTCAGAGC TGCCCTGATG ATCTACATCA TTTCGAGATTA
480 TTATCTGTTC CCGTGTGGTC TGCTGGGAG TAAATATGATA ACCCGCTTAC ACCCGCTTGT
540 AGCGTTTCTC ACTAGAGTAG CAGAGTGTGT TGGAAATAC TGTTCGATTT TAACTCTTTAA
600 CAAAGTGTTC GCTATGTGTC TTGATCCCTT CAGCTGAGAA GAGAGATGGA TACAGACCGCG
660 AGTAAAGAAA GCGATTTCTT CTTCCTACAT TAAATATGAA TTTCAGATGT TTTCCTTTTT
720 AAGAAAGAAA AGTCAATAGT TTAGATTTTT TTTCCTTTGA ATATGTGTGT TCTTCGACTT
780 TTGTGAGTAG TCTTTATAGA ATGAGAGTTT TCTAGACTTG TCAATTGAGC AAGAAATGCC
840 AGTTTATGAC AGGTATATGAC TAAAGAGAGC CCGTCTCAT CTGTACAGAA TGTGAAATGT
900 TTAGGAGATC CTGCGAGCTG TCACTGTGAA TTTCCTCTCT TATGTATTTT GCTATATATG

256

960 CCAAGCTGAA ATGCAAAATT TTCTAACAGC TTATATAGAG TTCTTTAGC ACTTATCTCT
1020 GTTTTATTC CAGTTTATGA ATATACTTAA TTATATATCC TCTATATATA ATGCAAAAT
1080 TTCTCTTTTA ATGCAATGCA ATATGTTTAA AAGAACAA CTAATTTGAA TTGCGTGGCC
1140 TTATCTGCC TTGTATTAAA ATGTCACTTGG TTGAGAGAA GTTGTATAT ATTAATCTACT
1200 TACTTGAAGC TGTATATTT TGTATACAT GTTTTATAAA TGTATCAGC AGACATGAT
1260 GATATGTTT GTTCTACAT TGTGTTTGT ATTATAGCA TTTCGATGAA AGGTATATTT
1320 AATAGCATTT AATATTTATG CTCTTTAGAA TGGAGACAG AAAACAAAC TTATATGTC
1380 TGAATATCT GAGCAATGAA CCGTGTGGC CTAGAGATTA TAACTCTAT AATGTCTCT
1440 TAAACACTT TTTCCTAATT AAAATCTTGG CAAATCTTGG TGTATCTTCC TGCTTTACAG
1500 CTACTGTTT GCTGTAGAGC AACCGAGCT GAGAGCTGAC TGTATATGA GTACATATAT
1529 CCGATTAAG CTAATTTTTC TCACTAAA

25 (2) INFORMATION FOR SEQ ID NO: 107:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2435 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

30 ATGAAAGATC GTTGGTGGGA AAAATGGCGG CAGCTTGGCG AACCTTGTGT CTGTGAGACA
60 GTTGGAGACA TGTTCGTGCG CTGTGGATGCG GTTCCAGATG TTACTCTCTT TTCTTTTGT
120 GGGGTCTGGG CAGGGGCGAC AACAAATGCG GAGCGGTGCA AGTTTCAGAT ACTTGAAAG
180 GAGACATGCG CTGTGAGAGC CCGTACAGCG TGTGGGACA GCGATTTCTT CAGTGTGAAA
240 TGTGATGGGC AATCCGATGCG TGAATGACCA GTATATGCGC CTTCACCGAG ATATGCAAG
300 TTAACAGAGT GCTTGTGGA ACCGGTGGCC ATGTTCCTTG AAGAGCTGCG AGTTTCAGAT
360 GCACTTGAAA ATCCATGAGC AAGGAAAGAA GAAATCTCAT GGGAGTGGCT TGGCAATGCG
420 GTTACAGAGG GATTCGATG CAGCCAGGCG CTGTATTTTGG GAAACATGGA CAATTTTGTG
480 GGGCTGGGAG TATTTTATGA CAGTACCC AATGAGAGAA AACAGAGAA GCGGTATATTC
540 CCGTACATCT CAGCGATGT GAGACAGGCG TGCCCTACAT ATGATCATGA GCGGATATGCG
600 GCGCTACAGAG AGCTGAGAGG CTGCAAGACC ATTTGTGCGA ATCTGATGTA CAGACACTTC
660 CTGTGATATC GCTACGTGCA GAGGCAATTA AACATATGTA TGGATATGTA TGGAGAGAT
720 GATGAGAGCG ACTGCAATGA AGTCCCGGGA GTTCCCTTCC CCGGCGATTA CTACTTGGCG
780

257

5
10
15
20
25
30
35
40
45
50
55
60

ACTCTCTCCA TCACCTGGGA TCCTCAGAT ATTCATGATG TCATTCTCT GAGTGTGTT
GACTGACAG TGGAGAGAAC CCCAGAGAG GAAAGCTCC ATCAGATAT GTTCTTGCC
TCAGTGACCA ATATGAGGT GCTTGAGATG ACAGCTCCAC TGGCCGCCCT GAGTGGCTG
GCCCTCTTC TCATCTCTCT TTTCTCCCTG GGTGTCTCT GTATTGCGA TAGTCTATG
TATCATACT TACACAAAT GCACGAGCA GAGCCGAAG CCTCTCTACT GAGCCCTCT
GCTGCCACCA CTTTCTGTAC TGTACCCAT GAGGTATGGA AGAGCAGGC ACTGGCTCA
GCATCAGCC TGGAGATGT TCTTCTCT AGCAGCTGCT TGGGACTAT ATTCTGTAC
TGGAGTTTG ATTCAGGGA CCCGCAATC CCATGGTTGT GCATGGGAC ATCTACTCT
GCTCTGGAA GCCACCCACC CCAGGCCAT GCTGCTGTGA TGTGCTTTC CCTGCACTCC
TTCCATGTGG GAGCAGAGT GTGAGAGAA TTATCTGTGT TGTGATGCA AAATCAGGA
ACAGATTTT ATAGCCAGG CTGCGTGT GTTTACTCA GAAGCCCTT CTACTCTAGT
TTTGAAATCA CAAGAAATTA AATAGTGA ACNCAGAG CTTTCTGAC ATCCATGCT
TGGTTTTC ATTGACCA ACCCTGACC TACTGTGGA GCTTCTTGG GAACAGGA
TGGAACTTC TTCCCTGCT TACTTCTCT TACTCTCAT CATTTGCTC TCTGTGCA
ACTGAGCTG GGAAGGCAT TTGATGCT CTCTGTGG GCTTGGGCT GCAGACACA
CCTGCTTTC ACTGCTTTC ATTAGTGG CCTAGGAGA TGGTTTCTG CTTTGATCA
CTGTTCCTTA GCATGGCTT TGGCTTAT GGCATGTCCA TGGCTTCCC ATTCAGTCT
CTTCAGGCC TCACTGAACT TTGGTAAG GTTGGTGA AAATCAGAG AGGCTGGA
GACATCTGG ATGCCATGA TTAGCTGAC ACTGACCG CTCCAGGTT GATCAACCA
AAGCAACAT TTGTCTATG GTCTGACAT GTGAGATGT TTCTGACTT CTGAGACCT
GCTTAGCTG ATGTGTTGTA GTTACGATT TTGAMTCC ACTTGAATG CTGAAGCTGT
AAGGAGGCTT TCTCTTACA CCTTGGCTT GATNTTGC CAGAGAGAA ATTTGGCTTT
TTTTTNTT ATGAGCAG AGCAGTTGC TGTCTCATG TTCCAGTCT GAGAGACAA
GACCTCAT ATCTGTGCT GCAGAGTTC ACTGTCAATG AGCAGCAGG CTTGATGCT
GGCTCTGTC AACCTTAT CTACTGCTT ATTGACAG GGTTCATG CTGCTCACT
TACTGGCTG GATTTAATC AGTTACAGC CAGATCTCC TTGAGGGCC TGGACTCTG
AGTCTCTCTA TGAACCTCTG TAGCTTAAT GAATTTCTTA AAATCAGGA TGGACCAA
AAAAAAAA AAAAAAAAAA AAAAAA AAAAA

258

(2) INFORMATION FOR SEQ ID NO: 108:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 805 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

5
10
15
20
25
30
35
40
45
50
55
60

ATCAAACTTA AGAATGAAT TCGAAGACT TCTCAGAGG ATTTGTATGT AACATGTTG
TATTCATTTT TGAAGAGTA ATTAAATTTG TAAAACTTCT GCTGTTTAC ACTGCATTT
CAATACAGCT AACTAATGAG AGGAGAGGG GAGGTCACTC TTTTGATGCT GCCCTGAAAC
CTCATTTGG TTCCCTGCTG CCTGCTTGG TGTGACCCAC GGAGATCCA CTCCAGCAT
GAGGTGCTC GTAGCTCTGC TGCCTACTCT GGTCTGTGGA TGCAGCGCGG TGAGGCTGG
GCTGTTTGA GAGGTCACTA ACCCTTCTCT GTTGTCTTGC CTTCTGCTGA AGACTGCGAG
AACCAACCG GGAAGCTGTC CTGAGGTGCC CTGCTGGGAG AGGCATATAG AATCTGTGAC
CTCTGACAC TGTGAAGCCA CCTGGGCTTA CAGAAACCAAC AGTCTTCCA GCATTTATTA
CAATCTTGA ATTCCTTGG GATTTTTC TCCCTTTTCA AAGCACTTAA GTCTAGATC
TACGTTTC CAGTCTCTGT CTGAGGTGAC TTAATAATC AGAACAAAC TTCTATTTC
CAGATCATG GGAGATACA CCTTTTCCAG GATNTATTT TTGGAAACA CTGAATGAA
ATCTTCCAG TATTAATAT TGTGTATTTA AAAAAAGAA ACTTTTCTGA ATGCTTACTG
GCCGTGATA CCAGCGATG TCCAGTTTA AAGATGAA AAAGATTA AACTTTTGA
GACAAAAA AAAAAAAAAA AAAAA

(2) INFORMATION FOR SEQ ID NO: 109:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1166 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

5
10
15
20
25
30
35
40
45
50
55
60

GCACGAGG GCGCAGCTG CAGGTCTCT GCTCAGGCT GAGAATGCG TCCCGCGCC
GGCTTCGGA GCATGCGGA CCCGAGAGC TGTTTTGA CAGACAGAA GCCCGGAAT
ACCTTCCAA CTCAGGATG ATTGATATCC AGACAGCAT GCTTGGCGA GCATGGAGC
TTCTTATCT CCCAGGAT AGCCCTGTT ACTCTGGA TATGCTGT GCACCTGGC
TGATGGAG TATCTGTCA GATGAGGCG ACTATGGGT GGGCTTGAT ATCAGCCCTG

360	CGATCTCTGA TGAAGCTCTG GATCGAGAGA TGAAGAGAGA CTTCTCTCTG GGGGATATAG
420	GGGAGGGAGT CCCATTCAG CAGAGAGAT TTGATGGTTG CATGAGAT TTCTGATGTC
480	AGTGGCTCTG TAAATCTAG AAGAGCTCTG AAAACCTCTG CAGAGCTCTG TACTGCTTTT
540	TTCTCTCTT TTTTCTCTT CTCTCCGAG GATCCGAGC TCTCTCTAG CTGAGAGCTG
600	AAGATCTAGA GAGATGAGAG CTGATCTAGA CCCAGGCTAC AAGAGCTAGC TTCTCCGAGT
660	GATGATGAT AGATGAGCTT AAGATCTAGA AAGAGAGAGA ATTCTAGCT TCTCTCTTTT
720	CTGGGCTTTC GATCTTTTAA CCGAGAGGAG TGAATGAAA TGAATGAAA GTTGAAGCCGA
780	GAGATCTCTT GTTACAGAT GAGAGCTTCC CATTTAGAT GTGAGAGGAG GAGATGATGA
840	GAGAGATCTG GAGATGAGT CTGAGAGAGA AAGAGAGGCA CAGAGAGCAG GCGAGAGAG
900	TGAGATCTGA GATCGAGTAC ACGAGAGGCA AGCGAGAGC CCCCTCTTAA GTGAGAGAGC
960	GCTCTCTGAA AAGCATCTGC CTCTGAGCTT TTCTTATATG TTCTGATCTAC AAGATGATAT
1020	TTTGAAGAG TTCTTAAATTT ATTTAAATTT TTCTCTGAGT AAAAAGAG TTCTCTGAGC
1080	CGAGCTGCT GCTCTGAGT TTTTATCTCA GATCTCTGAG AGCTGAGAT GAGAGATCTA
1140	TTTGAAGGCA GAGATGAGT AGCTCTCTG GATGATAT GAACTCTCT TTCTGAGGAG
1166	AAAAAAAA AAAAAAAAAA ACTCTGA

35 (2) INFORMATION FOR SEQ ID NO: 110:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 586 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

60	AAGAGAGAG AAGTGTGATA AAGAGAGAGC GATGATGAG GAGATCTAG TTCTCTCTCT
120	TTCTCTCTTA CTGAGAGAGC GGGTCTAGAG GAGAGCATTC CAGAGAGCAG GCGCTCTAGT
180	CGAGAGAGAG GTGAGAGAGC CAGAGAGCTT GAGAGAGCTT CCCATGATG AGCGAGAGC
240	GAACTCTAG TACGAGAGT AGCTTTCTT GAGAGAGGAA GTTGGCAGAG AATTGAGACA
300	ACTGAGCCCA GAGAGAGAGC AAGGCTCTCT GGGAGGAGAT GTGAGAGGCA TGAAGAGGAC
360	GAGAGAGAGC GAGAGAGTGG TTCTCTCTGAG CAGATCTCTG GCTGAGAGT CCGAGAGGCA
420	CGAGAGAGAG AAGAGAGAGT GCTGAGAGC GAGCTGAGAG AATTGAGACA CCGAGAGGCA
480	CGAGAGAGTGG GATTGAGAGG AGCTGAGGCA CCGATCTTAT GCTGAGAGTGG GAGCTCTCTGA

540	AAGATTTGAT GAGCTGAGAG ATGAGAGAGC TTAAGAGAG ATGCTGAGTTC GAGAGAGAGC
586	GCTTTCTCTG GTGAGAGAGC AAGATGAGGAA CTGATGAGCC ACTCTGA

5 (2) INFORMATION FOR SEQ ID NO: 111:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1134 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15 (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

60	ACCGATGAG CAGAGAGAG CAGATCTGGA AAGCTCTCTG GAGAGAGAGC GATCTCTGAC
120	ACTGAGCTCTG TTGATCTCTG ATCTGATAT CAGATTTCT GGGCTCTCTG GCTGATTTCT
180	ATCTATCTAG TTGAGAGTTT GCTTCTCTTC AGTCAATGAG CTCTTCAATC TACTCTCTTT
240	GCTCTCTATT TCGATCTCCA TGGTCAATGGA TGAAGAGTTC AAGAGAGTTT TGTCTCTGAC
300	AGGCTCTCTG CAGATCTGGA CTGAGATTC CATTTAGAGA ATCTGAGGCA ATTCTGAGTGC
360	CAGATTTAT TTCTGATTTA CTGAGATTC ATGAGATTC CCGCTTACAG CAGCATTTAT
420	GTGAGCTCTA AAGAGAGAGC GATCGAGCTC ATTCTGATTA TGTCTCTCTT GAGAGAGAAA
480	GAGAGAGAGT GATCTCTCTG TTGATCTCTA CAGAGAGTTT CAGAGAGTTT CATGATTTT
540	AAGAGAGTGA TTGATCTCTA CAGAGAGAG ACCCTCTGCA ATCTCTTTTGT GTCTGAGAAA
600	GATCTATCTAG GATGAGAGAG CAGAGAGT AGGCTCTCCA AAGTCTCTCT CAGCTCTGAC
660	GCTCTGAGAG GTCTATCTCTC ATCTATCTGA TACTGATCTAC GGGTTTCTGGA ATCTATCTCTG
720	TTATCTATCTA TTGAGAGGAA AAGAGAGGAA GTCAAGAGTTA TGAAGAGTTA GCGAGAGCTT
780	TGAAGCTCTT CTCTGATCTC CTGATCTCCA AAGAGAGTGC TTCTATCTGAA CAGATCTGAC
840	TGAAGATTTT TTGATTTTGA TTGATTTTGA TGAATCTCTC AAGAGATTTA TCAAGAGTGA
900	CAGATCTGCC CAGATCTCTAG AAGATGATTC TGAATCTCTCA GAAATCTCTA AAGAGCTCTT
960	ATCTCTGAGA AAGATGATTT GAGAGAGAT AGCGAGAGCT GTATCTCTAG GAGCTATGAT
1020	AAGAGAGAGC CAGAGAGAGC TGGCTCTGAA TTAAGATCTA AAGATCTGAT TTCTTATCTAT
1080	TTCTATCTATC AAGATGAGAG AAGAGATGGA GATCTCTGAT TTGGAGAGCAG GAAATCTCTT
1134	GTATTTTCTT AAGCATTTAA TTCTTACCCA GTTGTGATAG AAAAAAAAAA AAAA

60 (2) INFORMATION FOR SEQ ID NO: 112:

261

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1333 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

CACTTTAAG CTCCTGTGAG GAGATTGGA GCCCAGGCTT TCAGGGGACC TCTGCCCTTC 60
CTGCTCTCC TCACCTCTCC TCTCTTCTTG CAGGGCCCTGG GAAGGGCTTT GAGGGAGCTT 120
GGAGGCCATG TGAAGAGGGG CAGCCCTGGG CTGTCCCAACA GTTTAGATCC AGTTGGAGGT 180
TCTCCCTGGC TCTGTGAGGC CTGGGGGAT CTCTCCCAAC TTCAAGGCTTC GGGCAGCTGC 240
CTGCTCTCTT GTCTGTGCTT CAGCCCTGCA CAAAGCAGC TTGGTGACAC CACTCAGCCA 300
CCAGAGTAC GTGTATTACG GCTTTTCAGA TCACCTTCTT GTGGGGTGAA CGTAATGAGG 360
CGGGGCTGGT CATTGGAMTT TCCCTTGGAA AATGGTAAAC GACTCCATCC TTGACCCCGG 420
GATCAGCATG AAGGCAATTG CCAAGGCA GAGGCCAGG TGGTAGGAT TCCACCAAGG 480
CCAGAGGGGA AAGAGAGAA ACCCACCCTG TCTGTGCTGG GGGGCCCTGG GAGGGGTGTT 540
GAGTGAAGCC CTTCTCTACT TCTGTGCTTT TGTAAAGCT GTAGATATCC GCAATGGTTG 600
GCTGAGCCA GAACTCTCTT AATACATGG CTTCCTCCCC ACCCTTCTGT GGGGAGTCTT 660
TTTTAAAAA ATCTGTGGA TATAAATG GCTCTCTGCT GCTTCAAGCT ACCCTCTCCT 720
CTGCTGACTT AATGTGTGA TTCTGTCTCT TCAGATATTT AAGGCTGTGA GGTGTGTGA 780
GCCTTGAGT GTGTGTGTGT GTCCAGCA CTGTCCACTG TCCAGGAGAT GCATGTCTTT 840
GTATTGAGA TATTCTGTGA ACTCATCTC TTGGTGCTCA CGATTGCCAT GGCATAGGG 900
CCAGATGCC GTATCTGCTG CAGACATGAT TGTTCCTTGT TCTAGAGGTT TTCTTGTTTT 960
CGAATCTTGC CTGATGATTC CAGCCAGCC AAGGGGCCCTA GATTTCAGCT CTGTCTTGGG 1020
CTCTGGGCC AGGTGCAGGA ACATCTGAG CCACCTCTGCT GGCACCTCC AGTGGTGTCT 1080
GACCACAGGA TGGGCTTTGT TTACACTCAT TTTCACCTTG ATTCTTGCC CCACTTTCTT 1140
AAGAGAACT TCAAATATCT GAGGCTTGG AGATGAGAA ATCTAATCTT GGTGGGAC 1200
GGTGGCTCT GCTGTGATC CTAGCACTTT GGGAGGCTGA AGCTGAGGA TCACCTGAGC 1260
TCAGGAGTTG GAGACCAAC CTGGCAACT AACAGACC TGTCTCTACA AAAAAAAAA 1320
AAAAAAAACT CCA 1333

(2) INFORMATION FOR SEQ ID NO: 113:

(1) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1015 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

GGCAGAGCG GCACGAGCG CACGAGTGA CTTCAAGTGT CGATCTTTT CAGCCTACNT 60
CAGAGAGTG GAGGAGCGC CGCCACCCAC CCGTGGGCT CCAAGATGCC CTTTGGGGA 120
CTGATTTGC AATCCAGCAG TAGCTGGGC TTGGTACATG GGTCTGTCTT CTCAGCAGC 180
GGAGGCCGG TTGGCTGGGT AAGCCAGAC AGCACCGTCT GCTGGCTGA TCCCGACAG 240
AAGATGGCG TCGGACTCT GGCCTGTGA ACATACCAC TGTGGGCTT GACCTTCATC 300
ACAGACACA GCTGTGGC AGCGGCCAC GACTGCTTCC CGTCTGTCTT CACTATGAC 360
GCCCGCGCG GATGCTGAG CTTGGCGCG CCGCTGGAGC TTCTTAAGCA GAGCTCGAG 420
CTGTGCTTGA CGGCCCGGA GGCCTTCAG AACTGGACA AGAGGGGAG CTCGGGGGT 480
GGCAGGCTG CGGGCGGGG CCTAGACTG CTCACACAGA ACAGGCTCAG CCAGATCTCG 540
GTGCTCAGG CGGCCAAGC CAGTGTCTG CAGTTCTGCA CCACTGCCAT GATGTGGGCG 600
ATGAGTATCT GGAATGTGA GAGCTTGGAG TCAAGCTTGA AGGACCTCAA GATCAATGA 660
CCTGTGAGA ATATGTTGC TTCACTCTAG CTGTGGGA AGCGGGAGA GGGGTACGG 720
AGCTAATGG TTGCTTCTT GATGTTCTT GGGGTACCAA TAGGATTTCC CATAGGGCT 780
GCTCCCTCA AAGGGAGGG GACAGATGG GAGCTTTCTT TACCTATTC AAGAAATGCT 840
GCTTTTCTT TAATGCTTT CATTTATGA AAAAAAAAAA AATGCCCCC AAGCNCAT 900
GCTGTGATG AACTGCTTCA AATGTGGAG GTAAATAAT GCACTGTGT AAAAAAAAA 960
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAGNC 1015

(2) INFORMATION FOR SEQ ID NO: 114:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1076 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

50

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

GGCAGAGCG GAAGCCATG CTCACAGAC TCTTCTCTG CAGCCTTAA TCGTCTGTA 60
CGAATATTC CGGCTTAG AAGCCAGCG TTGGGTGTA CTTATATG TTCTTCTGA 120
CCTACTCTT GTTATCACT TCGGGTCA TCAATTTGG ATTGCTGA TCGGGTTGA 180

ACTATGAG CCCGCTTCA GCTCTTTC CCAATTTC CTTCAGAG AAGATCTG 240
GCTTCTCA AATCTGCT CTCTGGGTA GGGAGTCA AACCTTTC ATGAGAGAG 300
GAATGGAG GCGTCGGT GTTACTTA AATCCGCT CAGCTTCAG ACCGAACT 360
GGATTCCT CAGCGAGAG GCGGTACT GCGCTTCAG TCGTATTC CACTACAT 420
ATGTGAGC CAGGAGGCG GAGCAAGT CTGAGTTCA AGCCCCCG GAGGAAAT 480
AACCGGCT TGGAGAGCC GAGCGGAG TACATGAG TCGTGGCAT GATCTTCAG 540
ATGTGAGC TCACTTCA GCTAAGTGT TGTGCTGG TCGTCTCA CTGCTCTTC 600
ATCAGCTTTC CAGCTTCG GAGCTCGAG GAGAGAGAG AATATGAG TACCTTCAG 660
CTGTCACT CTGCGGTGT GATCTCAT TCGAGATC CTCAGCCAT GACCGCCCA 720
TGTATGAG AGCTAGAG GATCAATT TGGAGCTGT CTATCTCA GCGTGGCT 780
TTGCTCTCA AACCTCTG CTTCAGCT CAGCTTCAG TTGCTGAT GAGCGCTT 840
CGTCCGCG AGCTGATAG AGGAAAGT GCGTCTCT AGGAGAGC CTAGCTTAC 900
CGCTTCAG TCGTTCGCT TCGCTCTG TGGGGAGAT GCTGTCAAG TTCTAGGG 960
TATCTATT CTCTCTGT GAACTCTT GTTATGAG TTCTCTCT TGAAGAAA 1020
AAAAAATA AAAAGAGT GGGGGGGCG GAAACCAT TCGCGGATA GTGAGT 1076

(2) INFORMATION FOR SEQ ID NO: 115:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1487 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

CCGCTCTGA TAACTATGC ATCCCCCGG CCTGAGAG TTAGCAGG AGCTAGGG 60
CCGCTGCT CCTCTATCA CTCAGAGCT CTTCGCGGT GAGCCGAG CAGAGATCA 120
GATCAAGA CCACTAGGG GCGAGAGG TCGCGGTCT CCGAGACT TCGCTGCTG 180
TGTGTGCT CCGCGCGT TTCTATCT TATCTACT GGGCAGCG GAGAGGAGC 240
GGCGAGAG GAGAGAGCA CCGCTCTGT GCGCTTCG ACCCGCAGC 300
CCGCTGCT CTGAGAGC TGGTCGCG AGCGGCTTT CTACAGGT GCGATGCT 360
AATAGAGC CAGCTGAT GTGAGCTGT CCGAGCTA CAGTCAAG TACTTCACT 420
ACTGCTCA CCGCGAG AGTTCATAG CAGCACTCC CCGTGAAG TACTTCAGG 480
GCTCTTCT CTCTCTCT ATGAGGCA TCAAGAGT CATGGAGG AAGATGACT 540

ACTTCAGG CCGCTAGAG ATCCGCGCT TCGCGGCT GATGAGCTG GCGAGAGC 600
TGGTGTGC GATTAGCA GCGCTTCG TCGTGGCTG TCGTGTGCT AGCATCTG 660
TCACTGCT GCGATGAG GCGAGCTCA TGGTCCCC CAGAGAGC GAGGCTTGG 720
TTGAGGCT CAGAGCTTC TTGATGAG TGGCGAAG GCTGCAAT ATGCGATCC 780
AAGCTTCA CCGTTCGCG TCAAGCTCT GTCGAGAG CTTGCTGAG TTCTACAT 840
GGGTATGT GCTGTGAG GCGCGCTGG GGTGTGGCG TCGCGGTGT CTCTAGCC 900
TTCTCTGAG GCGAGCGCG CTGAGACT GATGAGCT GAGCGAAT GCTTCAGCA 960
GAGGATTT TCAAGCGCG AGCGAGAG CAGCGAGC CAGTGAAG TGAAGCATC 1020
AGAGCGCG GCGAGAGC GAGCGAGC GCTGAGAG GATCTAGC CAGAGAGT 1080
CAGAGCTG AGCGAGCA CAGTGGAT GAGCTTAG GATCTAGC TCACTAGAG 1140
AGAGAGCA CAGAGGTA CAGTGAAT CCACTTAT GAAATCTCA GAGAGAGCA 1200
ATCAAGAG TCAAGAGAG GATCTGAG TCGCGGACT GGGAGGGCG AGCTGGGCT 1260
GATGATGA CATATGGAG AGAGGCTGCT CTCTGGGTG ATGAGATGT TCTGATCA 1320
GATGAGTGT CTCAGCGCG TCGTGAAGT ACTGAGCC AGCTCACT AAGAGCTGAG 1380
ATTGTGAT TTACAGAT AAAAGAGCA AAGAGAGC AAAAAAATA AAAAAAATA 1440
AAAAAATA AATGATAT CAGCTTAT CAGTCTAG ACTTCA 1487

(2) INFORMATION FOR SEQ ID NO: 116:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1350 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

GCGAGAGT GCGAGCGT GCGCTCTG CTCTCAAT GCGCGCGCT GCGCGCTGT 60
GCTCTGAG CAGCGCGCG GCGAGACT CCGCATAG AGCGCTTCA GCAAGAGCA 120
GCGCGCGCG CCGTTCCT GAGTACCA GTCTTCTCA AAAATGAG AGAGATAT 180
AATCTCAT TTCTATAT TCAATTAT GAGATGAG ATGTGTTCA CATGTGAT 240
GAGTACAG GCTGTCTA TCGAAAAAT GAGATCTA CAGAGAGC TTAAAGCT 300
ATTAAAGAG ATGTGAAAA AGAGAACTT CCGTATGAG CAGATTGTT CCGGTAAAA 360
GATATATCT GAGATATAG TCGATCCT CAGACTGAG AAGAGCGAG CAGCATGAT 420

265

AAACATCTG GCTGTGTGG TGACATGAC CCATGTGATG TGTGTGAAT TGGAGCGAG 480
GTATGTGCA GAGGTCAAT AATTGGGTG AAGTCTTAG GCATATTGGC TATGATTGAC 540
5 GAAGGGAAA CCGACTGGAA AGTCATGCC ATTAAATGGG ATGATCTGGA TCGAGCCAA 600
TATATGATA TCAATCATGT CAAGCGGTG AAACCTGGCT ACTTAGAGC TACTGTGAC 660
10 TGGTTTAGA GTATTAAGT TCGTATGGA AACCCAGAA ATGAGTTTGC GTTTATGCA 720
GAATTTAAG ATAAAGACTT TCCCATGAT ATTATTAAAA GCACTCATGA CCATTTGAAA 780
GCATTAAGA CTAAAGAAC GAATGGAAA GGAATCAATT GCHTGAATAC AACTTTGTCT 840
15 GAGAGCCCT TCAATGTGTA TCGTATGCT GCGAGAGCA TTGTGATGC TTTACCACTA 900
CCCTGTGAAT CTGCTGCAC AGTACCAACA GACGTGGATA AGTGTATCCA TCACCAAGAA 960
AAGTAATGAG ATTCTCTGG AATACAGCT GATATTGCTA CATGTGTTC ATCTGATGT 1020
ATTAGAGTA AAGTAGTAG CTTTTCUAG CTTTAAATTT GTAGACTCA TCTAATCTAA 1080
GTAAATCTG CTGTACTAA TCCATATAC TCAGATATGT ATCCATCTAA AGCATTTTTC 1140
25 ATATCTCAC TAAGATACT TTTHAGCAT GCTTAAATAT CAAGCAATT GTCAATTTGA 1200
AGTCACTGT GATAGATGT CGAAGGGAG CACATATGCG ATGTATATGT TACCATATGT 1260
TAGAAATTA AATTATTTTG CTGAATAAAA AAAAATAAAA AACTSGGGG GGGGCCGGT 1320
30 CCCATTTGG CCTTTTGGG GGGGTTTTTA 1350

(2) INFORMATION FOR SEQ ID NO: 117:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2527 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

CTCTTCTAC CTTCGGCGG CAGAGAACCC CGGCTGCTCA GCGGCTCGG GGGTCATGGA 60
GATCCCGGG AGCTGTGTGA AGAAGTCAA GCTGAGCAAT AACCGGAGA ACTGGGAA 120
50 GCGAGAGCA ACCATGTCTA CTTACCAAGC CCATCATGTC ACAGAGACA AGAGAGTCA 180
GTTGTGGGG ACCAGAGGTG GCTTTGCTGG TTGCACAGTT TGGCTAACAG GCTTGTCTGG 240
AGCGGAAG ACTACTGTGA GCATGGCCTT GGAAGAGTAC CTGGTTTGTG ATGGTATTC 300
55 ATCTCACT CTGATGTG AGAATHTG TCAAGGTCTC AATTAATATC TTGCTTTAG 360
TCTCGAAG AGAGAGGA ATGTGCGAG CATCGCAGA GTTCTTAAC TGTTTGAGA 420
60 TCTGCGTGA GTGTGATCA CAGTTTTCAT ATCACTTAC ACTCAGATC GCATCAATGC 480

266

5 AAGCAAAAT CATGAGGTG CAACTTTACC GTTTTTCGAA GTATTGTG ATGCTCTCT 540
GCATTTTCT GACAGAGGG ATGTCAAGG ACTCTACAAA AAGCCCGGG CAGGAGAA 600
TAAAGTTTC ACTGGATCG ATTCTGAATA TGAANAAGCA GAGGCCCTG ACTTGTGCT 660
720 GAAACAGAC TCTGTGATG TAAATGACTG TGTCCAGCAA GTTGTGGAAC TTCTACAGA
10 ACGGATATT GTACCTGTG ATGCATCTTA TGAATTAATA GAACTATATG TCCAGAAA 780
TAACTTCAAT TTGGCAAAA CAGATGGGA AACATTACCA GCACTGAAA TTATTAAGT 840
GCATATCCAG TGGGTCCAGG TTTTGGCAGA AGGTTGGCCA ACCCATTTGA ATGGCTTTAT 900
15 GAGAGAGG GAGTACTTGC AGTGCCTTCA TTTTGAATGT CTTCGTGATG GAGGTGTCT 960
TAACTTGTCA GTACTATAG TTCTGACTGC GACTCATGAA GATAAAGAGA GCTTGCAGG 1020
20 CTGTACAGCA TTTGCTCTGA TGTATGAGG CCGCGTGTG GCCATCTTTC GCATTCAGA 1080
GTTTGTGAG CACAGAAAG AGGAGCCCTG TGCAGACAG TGGGAGAGA CATGCCAGAA 1140
CCACCCAT ATTAGATGG TGATGGAACA AGGAGATTGG CTGATTGGAG GAGATCTTCA 1200
25 AGTCTTGAT CGAGTTTATT GGAATGATGG TCTTGATCAG TATGTCTTGA CTCTACTGA 1260
GCTAAGCAG AAATTTAAG ATATGAATGC TGATCTGTC TTTGCAATTC AACTAGCAG 1320
30 CCCAGTCAAC ATGAGCATG CCTGTATAT GCAGATACC CATAAGCAAC TTCTAGAGAG 1380
GGCTTACCG CGCCCTGTC TCTCTCTCCA CCTCTGGGT GCGTGGACA AGGATGAGA 1440
1500 TGTCTCTTG ATGTGGCTA TGAGCAGCA TGTGTGATG TGGAGGAG GAGTCTGAA
35 TCTTGAGAG CAGTGTGTG CCATCTTCCC ATCTCCCATG ATGTATGCTG GACCACTGA 1560
GCTCAGTGG CATTCAGAG CAGCATGCT TGCAGAGCC AACTTTTACA TTGTGAGG 1620
40 AGACCTTCT GCATGCTTC ATCAGAAAC AGGAAAGAT CTTTATGAGC CAGTCAATGG 1680
TCCAAAGTG CTGAGATGG CCCCTGGTTT AATCACTTGG GAATATGTT CCTTTCGAGT 1740
45 TCCAGCTTAC AACAGAAAA AGAGCGTAT GCACTACTAT GACTCTGAC ACCATGAGA 1800
CTTTGAATTT ATTTCAGAA CACGAATGG CAATCTTCT CTGAGAGCC AGAAACCAAC 1860
TGAAGTTTC ATGGCTCCA AGGCTGGAC CGTGTGACA GAATACTTCA AATCTTGA 1920
50 GAAGCTTAG GCTGTTAAC CAGTCACTCC ACCTTTGACA CATTAAGT AACAGAGGG 1980
GACCAATAG TCTCTGTGG CATTTCTTGG TGTGTCTGT CTGACATGC TTCTAAAAA 2040
2100 CAGACATTT TCTTAACTT GCATCAGTT TGTCTGCTT TATGATGCT GTTTTGACA
55 AGCTTAAC ACATGATGT TTAATGATC TTTTCACTT ATTAATGTTA TATTTCTTCA
ATACATTT AAATGTCT TTTTATATTA TATTTATGCT TCTGTGCTAT GATTTTCTA 2220
60 AGCTGTATA TTAGTTGTA CAGTAGTAT TCATATTA TCTTCTTTT TTTTCCCTTA 2280

267

AAAAAAAA AAAAAAAAAA AAAAAAAAAA CTTGTTTGA GATTTTACA 2340
AAGCTTTGT AGAGTTTGA TTTTCTTCT AATTGAAAA TGAAGATGC TTCTTCTCT 2400
CTTTCAGTC AGCTATGTG CTTTCAGCT GTTATATCT TATGATCTGT 2460
GTAACTCTG AATGATCTC TTTACTCAT AATTATATT TTTTGTCTC TTAATAAAA 2520
AAAAAAA 2527

(2) INFORMATION FOR SEQ ID NO: 118:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1098 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

CGATCAGAG AGAACCGGA AGGAAATGG TTGGCGAGA CAGCAGGGG TTCAATTGC 60
TATATAAA CACTGCTGT AGGATTNKC TATGATCTT TGAATGAA AAAAACTCT 120
CTTGTGCCC CTTCAAGCC TATTGAGAT GACCAAGAG TATATGAGA TGTTCAGAG 180
CAGATGATA TTGAGAGCA CAGTCAGAT GAAATGGAG GATATATCC TCCACGACA 240
GATGATACA TTTATGANTG GATTGAGAG GAAATGCTG ATGATGTTT CCTGTCTCT 300
CTTAAGCAT TGGACATGG AGATGAGTT TACATGATG TGAATCTC TGAATTCCT 360
GTTTCATAG CAGATGAGG TGAAGAACT AATGTGGA AACCTAAGC AGAGAAAAG 420
GACTTAAAG AGCTAAAAA GAGAAAAA GAAAAAAG ACTTCAGGA AAAATTGAA 480
TATGATGTTG AATTATGAT CATTATTTA ACTAAGTTA GACTTCAT AACTCTAAA 540
AATGTGGAA CAGAGATCT AGAGTAAA CTTGTGAT CTTAGAGAT TTATCAAAAC 600
ACAGTGAAC CAAAGTTCT CTCAGAAAT GAAAGAGGA AATTGCTTA TGTCTTCTCG 660
AGTACTAGG CGAGCATGA TGGAGATC TATGATGTA TTGCTATCG CTCATCTAT 720
GAGATGATC AGAGTCAGC TTTGTGATC GTGCTGTCT GATTAGTGC CATTGAGAG 780
TCTGATTTT AATTGAGAT TATGTGGTA TGAAGAAAT TATGCGAR AACCTATAT 840
TATCATTTGT TATGAATCC CATTATCTT TACAAAGCT TTAAAGTTG AACCTAGAA 900
ATATCTCTC TGTATATG TATTCGAG AGACTAGCT AGTGAATCT AAGATATAT 960
AAATTTGCA TTTGCGCTT GGTCAATAT ATGAGAGAT TGAAGTACT TCTTTTAC 1020
CAGCATGAA TAACTCTCT TCAAAAAA AAAATAAAA AATCTGAGC 1080

268

GGGGGGGGG TACCCAT

1098

(2) INFORMATION FOR SEQ ID NO: 119:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1679 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

TGACCCAGC GTTCCGAGG GATCCCTAC GAGTGAAGC CTTTGTCCC GCGGAGCTT 60
CGGAACTCT TTGAGCCCC CGAGCCGCTT CCGGAGCCC GCGGTGAGG GCTGCATTC 120
CAGCGGAGG CCGGAGCCC GCGCCCGAG CCGCCCGCC CCGTTCAGG GCGCCCGAG 180
CGCGCCGATG GTGAGGTGA GTTCAACTC GCTCTGACC CAGAGGAGG CCAAGAGGA 240
CGAGCCGAG AGCGCGAGG AGGCGCTCAT GATCCCGCC GACGCGTTC GGTGAGCTG 300
CAGAGATCA GATGATGAG TACAGTTTG CAAAGAGGA GCTGTGTCTT GGTGCATTC 360
CTTGCATCA GATTTATTC TTGAGAGGT TATCTAGGA GAGCTATCT TGTCAAAATA 420
TTTTCAGCT CAGCAGATG AGCTGACTA CTGTGATTA AATGATCA AAGATGAT 480
CAGCTTAAT GAGCCCTCG CAGATCCCC AGCTCTCTC TACAGAGCA TTGAGAGAA 540
TATTAATAT TTTGAGAGG AAGAGTTGA ATTATCAT GTGCTGTTC CAGAGTTGC 600
AGATGATAT CTTCCGACA TTGTCATGA CTTTACAGG AATCTTAG CATTATTTGA 660
TCTTAAGCT GATGAGCTT ATGTATCCC TGTGAAGCT TCAATGTTA TGCAGCCAG 720
AAGCTTAGG GATTAATTA TTAACATCA GGTTCGAGC TATTCGCTC AGTCCATCT 780
GATTCATAG CAGATGCTA TTACTGATG CATTGAAAC ATTCATCAC TGGCTTCTT 840
TATTTATCA CTGTCTGAG AGAGAGAAC TTACAACTC CAGCGAGG AATCTATTA 900
AGATATAGG AAGCTGAG CAGCAATTC TTTCGATAT GCGATTTTG AAAAGAAAT 960
TGCCTGAGG ACTTAATTT GTTCTGAGC AGTCAGAAA AACATATTC AGAAAAATTA 1020
ATATCAGAC ATAAACCA CTTTACAT TTGTGAGTG ATTATTTT AAACTCTCT 1080
TTGATTAGG TACAGAGAG GCTTTTACTA TCTTTGATC TCAATTAATC AATTAAACC 1140
ATTACTTAA AATTTTCTC TTTCAGAGG TGTGTCTTT TATATTTGA TTGTAATCTG 1200
TATGAGTGA TACATATAG TACATCTAC CTTAGGAT AGAGAAAT ACATTTCTT 1260
TAAATCATT ATCTGATTT TTAATTTTA TTACATTT CAGAGAGG GATTTATAG 1320
AGATTAATCA TATATATCA TACSTAAA TGCATCAG TCACTATTT GATGTCTTA 1380

269

5 GTTCCCTGC TACCTAGTTT GTTAGTCAT TTGACACAC ATTAAATTT TCCTTAAT 1440
AAAATGCA GTATTTTCAG TGTCAATAT ATTAACTAT TTAGAGATG ATTTCCACT 1500
TTATGTTTAA ATNTCCTAGG CATCTGCTG ATATATATTT TAGAAATGT TTGGAATTA 1560
AGAAATAACT TGTGTACTA ATTGTATAA CCAATATCTG TGCATGGA TATAAATATC 1620
10 ACAAGTTGT TTAAWAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1679

15 (2) INFORMATION FOR SEQ ID NO: 120:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

25 TTGGCANCNG GGAGAGGAAA AGAGAGGAAA ATGGGGTTTG AGGACATGG CTTACCTTTC 60
CTGCTTTGA CCCATCAC CCATTTCTT CTCTTTTCC TCTCCCTGGT GCCAAAAAA 120
AAAAAAGG AAGCTTTAT CATGATCAA CAGGTTTCA GTCTTATCA AAGAGAGTG 180
TGGAAAGAGC TAAAGAAACC ACCCTTGTG TCCAACTCCA CTTTACCCT ATTATATCA 240
ACGAAACAC TGTCTTTTG GTTCCCTTTC TTACAGATGG ACCCTCTTGG AAGATATAC 300
35 GTATTCACG TTTTAGGCC TCAGTTTACC AAGATTAATA TATGTATATA TAACTTTAT 360
TATGTCTATA TCTTTGTGGA TAATACATTC AGTGTGTGCT GGTGATTTA TTATAATCTG 420
AACTTAGGTA TATCTTTTGG TCTTCCACAG TCATGTTGAG GTGGGCTCCC TGTATGGTA 480
AAAAAGCAGG TATATGTGAA CTTGACCCA GCTTTGTATC TAAAGCTCTTG ATATGTGATA 540
TACTCTTTTA AGTTTAGGCC CAATATAGG TAATGGAAAT TTCTTGCCCT CTGGGTTCCC 600
45 CATTTTACT ATTAAAGA CCAATGATTA TTTAATAATG CCACCAACT TGGCTTAGTT 660
AAGTAGAGT GTGACTGTG TGGCAGAGA GCTTCACACC TCATAGGTTG CAGAGAGCCC 720
AGGCTATG TTAATATCAT GCNCTTGAAA AGCAAACTT ATCTGCGAA GACAGACGA 780
AGCAATADAC GGTCACTCTG AATGATCCCT TTGAATTTT TTTTGTGTT GTTTGTTTAA 840
ATCAAGCCGT AGGCTGGTGA ACHTAGCTA CACACCCATA TTGTGTGTT TGTGAATGCT 900
55 AGCTCTCTG AATTGGATA TTGTTTATTT TTTATAGAT GTAAACCCAG TTTTATATC 960
TGCATGCGA ACAGTACT ATCTGTCTT AATATAACT GTTTACATTC ATTATGGGCT 1020
ATGTATGACC TTCAATTTCC AAGAAATAGA ACTTAGCTT AAGATATGAG ATGCTCTAAA 1080

270

ATGTCAGAT GGGAACTCT CTCGAAGTTC TCCAAACTC AGAGACAGA CTGCTTCTC 1140
CTAAATGAT ATTCTTTCT CCCTGTTTC TGGTATTTTC TAGGCATCTT TCTCACACAA 1200
5 GCCATAACC TTTTATCTT CAAATAGGCC GTATACTGG NGGAGNCTT GGTGGGTATA 1260
TAATACTGCT WCCACACAG GGGTTCTGGA TGTACACAG GTTAATCT 1308

(2) INFORMATION FOR SEQ ID NO: 121:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1411 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

15 GGACAGAG GACCCCGGA GAGAGAGGC CAGAGAGGC AAGGAGGA GTCTCCAGA 60
GACCCGGA CAGCATGCC CAGGCCCTG TTTCAGGCC TTTCAGATAT ATCCATCTCA 120
25 CAAGACATCC COTTAGAAG AGAATCAC ATTCCTATGA GATCTGCAT CCGGAGTTT 180
GACAGTCCA CATTAATGA ATCTGTGGC ATPACCATCA TCGCTGATCT AAGAGCTGT 240
30 GGGAAAAAT TCATCATGT TTGTACCCA AGGAAAGTA ATACTCTTT GAGAGATGG 300
GATTTGTGG GCCCTTGAT CTTTGTGTS ACATCGCAT TAATGCTGA AAGAGACTCT 360
GCAGATAGT AAAAAGATG AGGCCCCCA TTTCAGAGG TGTGTGCTAT TGTCTGTTT 420
35 GTTCAGTTA CACTACCT CAACTCAAA CTCTTGAG GACATATC TTTTTCAG 480
AGCTCTGTG TGTGGGTTA CTGTACTT CCTTGACAG TAGCAATCT GATTTGCCG 540
40 CTGCTACTT TGGCTGATCC AGGACCTGTA AACTTCATG TTGGGCTTT TGTGGTGT 600
GTGATGTTG CTTGTCTAT AGTGTCTCC ACAGCTTTC TTGCTGATAG CCACCTCCA 660
45 AACCCAGAG COTTAGCTGT TTAATCTGT TTCTGTTT ACTTTGCTAT CAGTTGATG 720
ATCTCACTT TTAATCTCA GTAAATCAG AATGGGAAT TAAAAACAG TGAATGAAA 780
GCACATCTA AAGATGCAAT TCACCATGA CTTTGTCTC TGGGCTTAT TGTCTAAT 840
50 TTGGAGTTA TTGATACTG AGTAGGTGAG GAGTTTAAA GGGAGCCATA TAGCACTGT 900
ACCCCTTAT TGAGGAAGT ATGTTGAAA GGCTGTCTT TTCTCTCTTA ATGTCATTT 960
TTTAAAAATA CATGTGCTA CTACACAG TATATAATCC CTCTTAAG CATCATGAG 1020
1080 TACCGTGTG CATTGGGT GACAACTGT GACTTGGGA GCACATAGT ACATCTTCA
AGTGTATAG AGTGTATAC TATTTCTGT TTGAGATA CAGTTTCAG TCCAGCTCT
60 AAGACATG CTTATGACT ATTAGATAT GCTCTCTTT TCATAATTA AATATACATG

271

1260 TCTATATCCA TTTTCTTTTA TTTCTCTCTC TTAACCTTAA AAAGCCATAG AAGAGAGTTA
1220 GAGATGGATT CATTACACGA GAAATGAGAA ACATGCAATTA ACCAAATATC AATATTTGAT
1380 CAGGGAAAT TCTATATCTG TTGCATAAAA AAAAAAATAA AATCTGAGG GGGGCGCGGT
1411 ACCCAATCCG NGATATGAT CGAAACAAAT C

(2) INFORMATION FOR SEQ ID NO: 122:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(*1) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

60 GATTTCGCTT TTTTTCACCG ACTGCGCGCC CTTCCGGAAG CTTTTCACAC TTTTCCAGAG
120 TTTTCTGGGA GGGGCAAGAG GGGATGAGGA CTGCAATATA TGAATCCCG GAGCAAGGGA
180 GGGGCTAAG AATGAAATCG TGTTCGAGCT CGAAGACGAG ATGCACATCC GGGGCTATAC
240 GAGCCGACG CAGGCCGACC GTGTGTACAG CAATACATTT CTGTGCAATG GGGTCCCTTC
300 TGTCTTACAG GCGCTTCTGC CTCTCTGAGG GGGGCTTGGC GCGCGAATG AGAATACTTG
360 CAGCCGACAG AACTCTTGGC ACTTCGAGGA AATTGAGCCC CAGAGCGGCC AGCTTTCGCG
420 AAGCCAGACC GAGCTGAGCT TCACTTTTAA CAGAGCGATG GCGAAGACG AAGCCATGGA
480 GAAATCTCTG GTGTATCTCG TGTATGTAGC CTCTGTCAGT GGGCTCTCTT GAGTGAAGCG
540 CAAAGCGACG AAGGAGCTGC AAGCCAGAGC AATGTATAGC GCGAGACGAC TGGCGAGAGA
600 GAGAGGTGAC GCGGCGCTCG GCGAGCTGCT GCGGTGATCT AAGCATGGA CAGCGCGTAA
660 CTTGACTCTG AAGTGTAGGA GCGGACTTTA CCGAGCCGAC TCAATGAGCT TCGCTGATGA
720 CTTTCTAGCG AAGCAAGCA GCACTGACAC TCGAGAGACT CGAATATCAA CTTCGCGCAC
780 AAGCGAGAG CCGTGCATCT CATTCAAGAG TGGGCGCGCG AAGCACCGA CCGCAAGCTG
840 CCGGAGATGA CGAAGAGCT GAGAGCGACG GAGCGCGCCC TGCATATCAA GCGCATTTTC
900 TTCAAGCCAC ACTGAGATGA GAAATTCAC GACAAAGATG TGGACAGCG TGGCTTCAATG
960 GTGACTGAGT CCAATATCTT GAGTGTCAAT ATGATGACAC GAGAGAGGCT GTACAACTAC
1020 TACAGACAGG AAGAGAAAA GCTGCAATC GTGAGATAGC CCGTGGCCCA CAGACTTTC
1080 AAGCTTATCA TCCATATGCC CCAATACATG GAGCTCTCG AGGCTTTGA AAGACTGTAA
1140 ACCAAGAGC AGCTAGAGAT CTGAGATGCG AAGATGAGA AAGAGGCTT TCCATATTC

272

1200 TTGCGCAAGG GTGTGTGTGA GTTGACCAAT GACTGTGAGA AACACTGGC TGGGCTGGCC
1260 CTGATGAGG CCAATGACAA GAAACAGGCC GACTTTTACG GATATGTACG CAAAGAGAC
1320 CTGTACCTCG GAGAGGTTT CAGAGCACG GCTTTTGAGT TGGACAGAGA TGGCAAGCCC
1380 TTGACACAG AACATCACG GCGAGAGAG CTGCGAAGCC CAGAGCTTTC TACGCGACCC
1440 AAGCTTCAT CTTCATATG GCGACACCC AAGCGGCTTC CCGCTATCTC ATTGGGCGCC
1500 TGTTCGCGCC TAAAGGTGAC AAGATGCGAG ACGAGTTATA GGGCTGTAGG GTGCACAGAG
1560 GATGCGAGGA GGCATTCAAA GCTTCTGAG ACACATGAGT GCTATGAGGG TTGGGCGGGA
1620 GATGAGGTAC CAGCTTTGGA TACTTCATCG GGTGGGCGTG GAAATGACGA CCGGGCTTCC
1680 CCGTGGCTCG AAGCAAGCTT CCGACATAGA ATTCATCTCA CTTGACATAG GCGCCGAGAT
1740 ACGATATCC TGAAGCCGGA AACTCCAGT CCGTGTAGAC CTGGGCCATA GTCAATCTGC
1800 CTGCGCTGAA AATCCCAAGT CAGGCTTGGC TCAATCAGTA TTCAATTTTA TGAACAGTTA
1860 CTTTCTACG TGTGAGACAA AATGTAGCTA GGGGGTCAAG CCAAGCTCTT TCTGACACTA
1920 AAACATCTCA GCTGCTTCCC CAGCTTATC CCAACTCTTC CCACTGTAAA AACTAGTTC
1980 TCGAGCTCCT GGAACGAGCC ACGCCAGAAA TCACTTGGCC GCAATGAGCC GATATTGAAA
2040 GAGACTTCCA GAGAGGAGCT CTGGGACAGC TGTGGTCAAG AACGATGATG TGTGAGCTTG
2100 TGGGATGAAA CTTTCTTCTT TGTTCCTTCC TTTTGTAGTT CTTCAGAAAT AAGAGAGAAA
2160 GGGGAAAT GAGCTTTTGT TGTATCAT TCAAGAACTT AATTGTACAT TTTTTTTTTC
2220 AATTAATCTT TTGCATGAC AAAAABAAAA AAAAAAGGAG GAGCGCTTCC
2256 TGAAGGATTC CTTCGAGAG AGCCCAATCG AAAATN

(2) INFORMATION FOR SEQ ID NO: 123:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 829 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(*1) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

60 ATTCGCTCCC TCCCTGCTTT CAGGCTCTTC TGCCTCTCG AGCGGCGGCT GCGCGCGGAG
120 GTGAAGAAAC CTGAGAGCCG AGCACTCTCT GCACTCTCGG AAGATTTAGG CCGCAAGCG
180 GCGACCTGTT CCGAGCGGAA GCGGCTTGGC CTTCAGCTTG TACCAAGCCA TGGCCAAAGGA
240 CAGGCGATG GAGAACATCC TGTGTTCACG CTTGATGTGT GCGTCTGAGC TGGGAGCTGT
300 GTGCTTGGCC GCGAAGCGGA CAGCGGCTC GAGAGCGAG GAGATCTGA GCGCCGAGGA

273

360 GCTGCGGAC GAGAGGTGC ACCCGGCTT GGGGAGCTG CTGGCTGAC TCAGCACTC
5 CACGGGCGC AGCTGACTT GGAAGCTGG CAGCGACTG TACGGAGCCA GCTCAGTAG
420 CTTCGCTGAT GACTTGTGC CGACAGCAA CGAGCACTAC AACTCGAGC ACTCCAGAT
480 CAACTTTCG GAGAGGCGA GCGCGCTCA GTCCATCAAC GAGTGGGCG CGCAGAGCCAC
540 CGAGCGGAG CTGCGCGAGG TCACAGAGA GGTGGAGGCG ACGAGAGCGG CCTGTGAT
600 CAGCGCATG TTCTTCAGC CACACTGGG TAGAATTC CACACAGA TGTGGACAA
660 CCGTGGCTTC ATGTGACTC GTTCTATAC GTTGGGTGC ATGATGATGC ACCCGAGG
720 CCTTACAC TACTACGAGC ACGAGAGGA AAGCTGCGA ATCTGGAGA TCGCCCTGGC
780 CCAGAGCTC TCACGCTCA TCATCTCAT GCCCATCAC GTGGAGCT
829

(2) INFORMATION FOR SEQ ID NO: 124:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2221 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

35 CTTCCGAG GATTTCAC TTTCAGAG TTCTCGGA GGGGAGG GGGTGGGA
120 CTGCGATATA TAGATCCCG GACAGGGGA GCGGGTAG ATGAGATG TGTCGGCT
180 CGAGAGGAG AATCACTTC CGCGCTAGC CAGCGCCAGC CAGGCTCAG
240 CAAACACTT CTGCGCATG CGCTCTCTC TCTTCTCAG CGCTTCTGC CTCTGGAG
300 CGCGCTGCG CGCGAGGTC AAGAACTG CAGCGGAGC AGCTCTGCG ACTCGGAGA
360 AGTTGAGGCC CAGGCGGCG ACCTTTCG AGCGGAGCG CCTTGGCTT CAGCTTGATC
420 CAGGCGATG CGAGAGACA GGCAGTGGAG AACTCTCTG TGTACCGCT GTGTGTGGC
480 TGTGTGCTG GGTCTGTG CTGTGGGCG AAGCGAGCA CGGGTGGCA GGTCAAGCCA
540 GTGCTGAGCG CCGAGAGCT GCGCGAGAG GAGGTGACG CGCGCTTGG CGAGTCTG
600 CGCTGACTCA GCACTGCA GGGCGGCA GTCAGTGA AGTGGGAG CCGACTGTAC
660 GAGCGAGCT CAGTGAGCTT CGCTGATG TCTGTGGCA CAGCAGAG CACTGACCT
720 GCGAGACTC CAGATCAAC TTCCGACA AGCGAGCG GTGAGTCA TCAGAGATG
780 GCGCGGCGC ACCAGCGAG CGAGCTGCC CAGGTACAC AAGAGGTGG AGCGCAGGA
840 CGCGGCGCTG YTAGTCAGC CAAATCTT CAGGCGAG TGGATGAGA AATTCAGCA

274

900 CAGATGGTG GACAACTGT GCTTCATGT GACTGGTCC TATACGTGG GTCTCATGAT
960 GATCAACCG ACAGGCTCT ACUACTACTA CGAGAGAG AAGGAAAG TCGAATCTG
1020 GGAGATGCCC TTGGCCGACA AGCTTCTCAG CCTATCTATC CTATGCCCC ATCAGTGGG
1080 GCGCTTCAG CGCTTTCAAA AGCTGCTTAC CAAGAGCAG CTGAGATCT GATGGGGAA
1140 GATCCAGAG AAGCTGTTG CCATCTCTT GCCCAGGCT GTGTGGAGG TGAGCCATGA
1200 CTTCCAGAA CAGCTGGCT GCGTGGGCT GACTGAGGCC ATTGACAGA ACAAGGCGGA
1260 CTTTCAGCG ATGTAGGGA AGAGAGCT GTACTGGCC AGGTGTTC ACCCCAGCG
1320 CTTTGAATT GACACAGAT GCACCCCTT TGACAGGAC ATCTACGGC GCGAGAGCT
1380 GCGCAGCCA AGCTGTCTTA CGCGAGCAC CCTTCACT TCTTAGTGG GACACCCAA
1440 AGCGGCTCC TCTATTCTAT TGGGGCTG GTCCGGCTA AGGCTGACAA GATCGGAGC
1500 GAGTTATAG GCTCAGGCT GCACACAGA TGGCAGGAG CATCCAAAG CTCCTGAGC
1560 ACATGGTGC TATTTGGGTT GGGGGGAGG TGAAGTACA GCTTTGATA CTCATGGGG
1620 TGGGGTGGG AATCAGACC GGGGTTCGG TTGCGCTGAG CGAGCTTTC CAGCTGAGT
1680 TCTCTCACT TGGACATGG CCGCAGTATC CATGATCTG AGCGCGGAA CTCACATCC
1740 TGTGGAGCT GCGCATAGT CATCTGCTT GCGCTGAAG TCCAGATCA AGCTGCGCT
1800 AATCAGTAT CATATTATA GCGAGTACC TTCTGACTG TGAGACCAA TTGAGCTAGG
1860 GGGGTGAGC AGCCCTCTTC TGACACTTAA ACACCTCAGC TGCTTCCCA GCTCTATCC
1920 AACCTTCC ACTATATAAA CTAGTGTCTG CAGCGGCTGG GACAGGAGC CCCAGAAATG
1980 ACTTGGCGC AGTGGGCGG ATTGAGAGG AGCTCCGAG AGGGGCTTCT GGGGAGACTC
2040 TGGTCAGAA GCACTGTGTC TGGGTGTG GGGATGACT TTCTGTTTG TTCTTCTTT
2100 TTTTATCTT TCAAGATAG GCGAGGAGG GCGAATCA GCTTTGTG CTATCAATCC
2160 AAGAATTAT TTGTACATTT TTTTITCAA TAAACTTTT CAAATGACA AAAAAAAAA
2220 AAAAAAAAA MMGGGGGGGG GCGGCTCTTA GAGGATCCC TCGAGGAG CCCCATCGA
2223 AAT

(2) INFORMATION FOR SEQ ID NO: 125:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

Met Lys Lys Gln Ser Lys Arg Cys Leu Trp Lys Pro Gly Ser Leu

275

1 5 10 15
Arg Arg Leu Trp Trp Met Arg Ala Leu Leu Ile Leu Lys Tyr Ile
20 25 30

(2) INFORMATION FOR SEQ ID NO: 126:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

15 Met Lys Lys Ser Leu Glu Asn Leu Asn Arg Leu Glu Val Met Leu Leu
1 5 10 15

20 His Leu Thr Ala Ala Phe Leu Glu Arg Ala His Ile Leu Thr Thr
20 25 30

Arg Met Ser Leu Gly Phe Glu Ser Pro His Leu Thr Met
35 40 45

(2) INFORMATION FOR SEQ ID NO: 127:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

35 Met His Asn Glu Arg Glu Val Phe Leu Phe His Leu Ser Asn Tyr
1 5 10 15

40 Leu Leu Ser Ile Asn Ser Val Pro Gly Thr Leu Leu Ala Ala Thr Tyr
20 25 30

Cys Leu Asn Met Thr Tyr Gly
35

(2) INFORMATION FOR SEQ ID NO: 128:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

55 Met Arg Lys Lys Phe Leu Leu Ala Glu Val Phe Leu Ser Leu Ser Val
1 5 10 15

Met Pro Ser Met Pro Val Thr
20

60 1

276

(2) INFORMATION FOR SEQ ID NO: 129:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 110 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

10 Met Val Leu Leu Cys Leu Leu Leu Val Pro Leu Leu Ser Leu Phe
1 5 10 15

Val Leu Gly Leu Phe Leu Trp Phe Leu Lys Arg Glu Arg Glu Glu
20 25 30

15 Tyr Ile Glu Glu Lys Lys Arg Val Asp Ile Cys Arg Glu Thr Pro Asn
35 40 45

20 Ile Cys Pro His Ser Gly Glu Asn Thr Glu Tyr Asp Thr Ile Pro His
50 55 60

Thr Asn Arg Thr Ile Leu Lys Glu Asp Pro Ala Asn Thr Val Tyr Ser
65 70 75 80

25 Thr Val Glu Ile Pro Lys Lys Met Glu Asn Pro His Ser Leu Leu Thr
85 90 95

Met Pro Asp Thr Pro Arg Leu Phe Ala Tyr Glu Asn Val Ile
100 105 110

(2) INFORMATION FOR SEQ ID NO: 130:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 63 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

40 Met Leu Leu Leu Phe Ile Tyr Phe Tyr Ser His Pro Ala Pro Val Pro
1 5 10 15

45 Ala Gly Ala Thr Ser Lys Pro Arg Tyr Arg Val Ile Thr Cys Gly Pro
20 25 30

Ala Ser Val Phe Ser Thr Ser Phe Ser His Ser Pro Pro Ala Arg Cys
35 40 45

50 Leu Gly Arg Leu Glu Glu Met Phe His Phe Gly Leu Ala Ser Gly
50 55 60

(2) INFORMATION FOR SEQ ID NO: 131:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

60

277

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

Met Pro Phe Pro Ile Ser Ile Leu Gln Leu Cys Leu Gln Ile Ser Asn 15
 1 5 10
 Leu Ser Phe Cys Leu Gln Lys Ile Tyr Lys Ile Pro Phe Val 30
 20 25

10

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 53 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

Met Ala Ala Ala Cys Arg Ser Val Lys Gly Leu Val Ala Val Ile Thr 15
 1 5 10
 Gly Gly Ala Ser Gly Leu Gly Leu Ala Thr Ala Asp Asp Leu Trp Gly 30
 20 25

25

Arg Glu Pro Leu Leu Cys Phe Trp Thr Cys Pro Thr Arg Val Gly Arg 45
 35 40

Pro Lys Pro Arg Ser 50

30

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 57 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

Met Leu Leu Val Tyr Asp Leu Tyr Leu Xaa Pro Lys Leu Trp Ala Leu 15
 1 5 10
 Ala Thr Pro Gln Lys Asn Gly Lys Gly Ala Arg Xaa Gly Asp Gly Thr 30
 20 25

45

Pro Ala Gln Ala Phe Trp Asp Phe Trp Ser His Leu Ile Ser Ala Asp 45
 35 40

Pro Gln Thr Trp Glu Arg Ala Ala Pro 55
 50

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 216 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

55

60

278

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

Met Arg Leu Ser Ala Leu Leu Ala Ser Lys Val Thr Leu Pro 16
 1 5 10
 Pro His Tyr Arg Tyr Gly Met Ser Pro Pro Gly Ser Val Ala Asp Lys 30
 20 25

5

Arg Lys Asn Pro Pro Trp Ile Arg Arg Arg Pro Val Val Glu Pro. 45
 35 40

10

Ile Ser Asp Glu Asp Trp Tyr Leu Phe Cys Gly Asp Thr Val Glu Ile 60
 50 55

Leu Glu Gly Lys Asp Ala Gly Lys Gln Gly Lys Val Val Gln Val Ile 80
 65 70 75

15

Arg Gln Arg Asn Trp Val Val Gly Gly Leu Asn Thr His Tyr Arg 95
 85 90

20

Tyr Ile Gly Lys Thr Met Asp Tyr Arg Gly Thr Met Ile Pro Ser Glu 110
 100 105

Ala Pro Leu Leu His Arg Gln Val Lys Leu Val Asp Pro Met Asp Arg 125
 115 120

25

Lys Pro Thr Glu Ile Glu Trp Arg Phe Thr Glu Ala Gly Glu Arg Val 140
 130 135 140

Arg Val Ser Thr Arg Ser Gly Arg Ile Ile Pro Lys Pro Glu Phe Pro 160
 145 150 155

30

Arg Ala Asp Gly Ile Val Pro Glu Thr Trp Ile Asp Gly Pro Lys Asp 175
 165 170

35

Thr Ser Val Glu Asp Ala Leu Glu Arg Thr Tyr Val Pro Cys Leu Lys 190
 180 185

Thr Leu Gln Glu Glu Val Met Glu Ala Met Gly Ile Lys Glu Thr Arg 205
 195 200

40

Lys Tyr Lys Lys Val Tyr Trp Tyr 215
 210

45

(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 49 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

Met Ser Leu Arg Arg Gln Lys Ser Ser Phe Arg Leu Met Val Met Ser Leu 15
 1 5 10

55

Thr Ile Leu Lys Leu Ser Lys Thr Thr Val Leu Cys Leu Arg Cys Leu 30
 20 25

60

279

His Ser Leu Lys Leu Thr Trp Arg Asp Gly Ala Arg Cys Ile Asn Ala
35 40 45

5

Glu

(2) INFORMATION FOR SEQ ID NO: 116:

10

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 68 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

Met Ser Gly Ser Phe Ile Leu Cys Leu Ala Leu Val Thr Arg Trp Ser
1 5 10 15

Pro Gln Ala Ser Ser Val Pro Leu Ala Val Tyr Glu Ser Lys Thr Arg
20 25 30

Lys Ser Tyr Arg Ser Gln Arg Asp Arg Asp Gly Lys Asp Arg Ser Gln
35 40 45

Gly Met Gly Leu Ser Leu Leu Val Glu Thr Arg Lys Leu Leu Ser
50 55 60

Ala Asn Gln Gly
65

(2) INFORMATION FOR SEQ ID NO: 137:

35

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

Met Cys Phe Arg Phe Phe Leu Phe Cys Ser Arg Ile Leu Leu Lys Leu
1 5 10 15

Phe Phe Leu Leu Phe Pro Ala Ser Ala Phe Pro Leu Ser Thr Arg Ser
20 25 30

Ser Leu Ser Val Asn Glu His Val Val Ser Pro Arg Ser Thr Val
35 40 45

Ser Ile Ser Arg
50

(2) INFORMATION FOR SEQ ID NO: 138:

55

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 541 amino acids

(B) TYPE: amino acid

60

280

(D) TOPOLOGY: linear
(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

Met Val Arg Thr Asp Gly His Thr Leu Ser Glu Lys Arg Asn Tyr Gln
1 5 10 15

Val Thr Asn Ser Met Phe Gly Ala Ser Arg Lys Lys Phe Val Glu Gly
20 25 30

Val Asp Ser Asp Tyr His Asp Glu Asn Met Tyr Tyr Ser Gln Ser Ser
35 40 45

Met Phe Pro His Arg Ser Ser Glu Lys Asp Met Leu Ala Ser Pro Ser Thr
50 55 60

Ser Gly Gln Leu Ser Ser Gln Phe Gly Ala Ser Leu Tyr Gly Gln Gln Ser
65 70 75 80

Ala Leu Gly Leu Pro Met Arg Gly Met Ser Asn Asn Thr Pro Gln Leu
85 90 95

Asn Arg Ser Leu Ser Gln Gly Thr Gln Leu Pro Ser His Val Thr Pro
100 105 110

Thr Thr Gly Val Pro Thr Met Ser Leu His Thr Pro Pro Ser Pro Ser
115 120 125

Arg Gly Ile Leu Pro Met Asn Pro Xaa Asn Met Met Asn His Ser Gln
130 135 140

Val Gly Gln Gly Ile Gly Ile Pro Ser Arg Thr Asn Ser Met Ser Ser
145 150 155 160

Ser Gly Leu Gly Ser Pro Asn Arg Ser Ser Pro Ser Ile Ile Cys Met
165 170 175

Pro Lys Gln Gln Pro Ser Arg Gln Pro Phe Thr Val Asn Ser Met Ser
180 185 190

Gly Phe Gly Met Asn Arg Asn Gln Ala Phe Gly Met Asn Asn Ser Leu
195 200 205

Ser Ser Asn Ile Phe Asn Gly Thr Asp Gly Ser Glu Asn Val Thr Gly
210 215 220

Leu Asp Leu Ser Asp Phe Pro Ala Leu Ala Asp Arg Asn Arg Glu
225 230 235 240

Gly Ser Gly Asn Pro Thr Pro Leu Ile Asn Pro Leu Ala Gly Arg Ala
245 250 255

Pro Tyr Val Gly Met Val Thr Lys Pro Ala Asn Glu Gln Ser Gln Asp
260 265 270

Phe Ser Ile His Asn Glu Asp Phe Pro Ala Leu Pro Gly Ser Ser Tyr
275 280 285

Lys Asp Pro Thr Ser Ser Asn Asp Asp Ser Lys Ser Asn Leu Asn Thr
290 295 300

60

281

Ser Gly Lys Thr Thr Ser Ser Thr Asp Gly Pro Lys Phe Pro Gly Asp
305 310 315

Lys Ser Ser Thr Thr Gln Asn Asn Gln Lys Lys Gly Ile Gln
325 330 335

Val Leu Pro Asp Gly Arg Val Thr Asn Ile Pro Gln Gly Met Val Thr
340 345 350

10 Asp Gln Phe Gly Met Ile Gly Leu Leu Thr Phe Ile Arg Ala Ala Glu
355 360 365

Thr Asp Pro Gly Met Val His Leu Ala Leu Gly Ser Asp Leu Thr Thr
370 375 380

15 Leu Gly Leu Asn Leu Asn Ser Pro Glu Asn Leu Tyr Pro Lys Phe Ala
385 390 395 400

Ser Pro Trp Ala Ser Ser Pro Cys Arg Pro Gln Asp Ile Asp Phe His
405 410 415

20 Val Pro Ser Glu Tyr Leu Thr Asn Ile His Ile Arg Asp Lys Leu Ala
420 425 430

25 Ala Ile Lys Leu Gly Arg Tyr Gly Glu Asp Leu Leu Phe Tyr Leu Tyr
435 440 445

Tyr Met Asn Gly Gly Asp Val Leu Gln Leu Leu Ala Val Glu Leu
450 455 460

30 Phe Asn Arg Asp Trp Arg Tyr His Lys Glu Glu Arg Val Trp Ile Thr
465 470 475 480

35 Arg Ala Pro Gly Met Glu Pro Thr Met Lys Thr Asn Thr Tyr Glu Arg
485 490 495

Gly Thr Tyr Tyr Phe Phe Asp Cys Leu Asn Trp Arg Lys Val Ala Lys
500 505 510

40 Glu Phe His Leu Glu Tyr Asp Lys Leu Glu Glu Arg Pro His Leu Pro
515 520 525

Ser Thr Phe Asn Tyr Asn Pro Ala Gln Gln Ala Phe Xaa
530 535 540

45

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 58 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

Met Ile Cys Pro Gln Cys Pro Leu Ser Leu Leu Cys Leu Ile Ser Ser
1 5 10 15

Leu Cys Ser Leu Val Ile Gln Ile Ser Leu Lys Thr Ile Arg Asp Ile
20 25 30

282

Thr Leu Leu Asn Met Val Gly Ile Lys Phe Ser Ile Ser Leu Ser Asn
35 40 45

5 Lys Ile Asn Ile Asn Ser Arg Thr Trp Xaa
50 55

10 (2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 202 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

Met Thr Leu Arg Pro Ser Leu Leu Pro Leu His Leu Leu Leu Leu
1 5 10 15

20 Leu Leu Ser Ala Ala Val Cys Arg Ala Glu Ala Gly Leu Glu Thr Glu
20 25 30

25 Ser Pro Val Arg Thr Leu Gln Val Glu Thr Leu Val Glu Pro Pro Glu
35 40 45

Pro Cys Ala Glu Pro Ala Ala Phe Gly Asp Thr Leu His Ile His Tyr
50 55 60

30 Thr Gly Ser Leu Val Asp Gly Arg Ile Ile Asp Thr Ser Leu Thr Arg
65 70 75 80

Asp Pro Leu Val Ile Glu Leu Gly Gln Lys Gln Val Ile Pro Gly Leu
85 90 95

35 Glu Gln Ser Leu Leu Asp Met Cys Val Gly Glu Lys Arg Ala Ile
100 105 110

Ile Pro Ser His Leu Ala Tyr Gly Lys Arg Gly Phe Pro Pro Ser Val
115 120 125

Pro Ala Asp Ala Val Val Gln Tyr Asp Val Glu Leu Ile Ala Leu Ile
130 135 140

45 Arg Ala Asn Tyr Trp Leu Lys Leu Val Lys Gly Ile Leu Pro Leu Val
145 150 155 160

Gly Met Ala Met Val Pro Ala Leu Leu Gly Leu Ile Gly Tyr His Leu
165 170 175

50 Tyr Arg Lys Ala Asn Arg Pro Lys Val Ser Lys Lys Lys Leu Lys Glu
180 185 190

55 Glu Lys Arg Asn Lys Ser Lys Lys Xaa
195 200

(2) INFORMATION FOR SEQ ID NO: 141:

60

283

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 217 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

5 Met Phe Leu Arg Leu Tyr Leu Ile Ala Arg Val Met Leu Leu His Ser
1 5 10 15
10 Lys Leu Phe Thr Asp Ala Ser Ser Arg Ser Ile Gly Ala Leu Asn Lys
20 25 30
15 Ile Asn Phe Asn Thr Arg Phe Val Met Lys Thr Leu Met Thr Ile Cys
35 40 45
Pro Gly Thr Val Leu Leu Val Phe Ser Ile Ser Leu Trp Ile Ile Ala
50 55 60
20 Ala Trp Thr Val Arg Val Cys Glu Ser Pro Glu Ser Pro Ala Glu Pro
65 70 75 80
Ser Gly Ser Ser Leu Pro Ala Trp Tyr His Asp Glu Asn Val Thr
85 90 95
25 Ser Asn Phe Leu Gly Ala Met Trp Leu Ile Ser Ile Thr Phe Leu Ser
100 105 110
30 Ile Gly Tyr Gly Asp Met Val Pro His Thr Tyr Cys Gly Lys Gly Val
115 120 125
Cys Leu Leu Thr Gly Ile Met Gly Ala Gly Cys Thr Ala Leu Val Val
130 135 140
35 Ala Val Val Ala Arg Lys Leu Glu Leu Thr Lys Ala Glu Lys His Val
145 150 155 160
His Asn Phe Met Met Asp Thr Glu Leu Thr Lys Arg Ile Lys Asn Ala
165 170 175
40 Ala Ala Asn Val Leu Arg Glu Thr Trp Leu Ile Tyr Lys His Thr Lys
180 185 190
Leu Leu Lys Lys Ile Asp His Ala Lys Val Arg Lys His Glu Arg Lys
195 200 205
45 Phe Leu Pro Ser Tyr Pro Pro Val Xaa
210 215

50 (2) INFORMATION FOR SEQ ID NO: 142:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 102 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

60 Met Ser Asn Thr Thr Val Pro Asn Ala Pro Glu Ala Asn Ser Asp Ser
1 5 10 15

284

Met Val Gly Tyr Val Leu Gly Pro Phe Phe Leu Ile Thr Leu Val Gly

20 25 30

5 Val Val Val Ala Val Val Met Tyr Val Glu Lys Lys Arg Val Asp
35 40 45

Arg Leu Arg His His Leu Leu Pro Met Tyr Ser Tyr Asp Pro Ala Glu
50 55 60

10 Glu Leu His Glu Ala Glu Glu Glu Leu Leu Ser Asp Met Gly Asp Pro
65 70 75 80

15 Lys Val Val His Gly Trp Glu Ser Gly Tyr Glu His Lys Arg Met Pro
85 90 95

Leu Leu Asp Val Lys Thr
100

20 (2) INFORMATION FOR SEQ ID NO: 143:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 112 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

30 Met Arg Glu Cys Glu Glu Ser Phe Trp Lys Arg Ala Leu Pro Phe
1 5 10 15
Ser Leu Val Ser Met Leu Val Thr Glu Gly Leu Val Tyr Glu Gly Tyr
20 25 30
35 Leu Ala Ala Asn Ser Arg Phe Gly Ser Leu Pro Lys Val Ala Leu Ala
35 40 45
Gly Leu Leu Gly Phe Gly Leu Gly Lys Val Ser Tyr Ile Gly Val Cys
50 55 60
40 Glu Ser Lys Phe His Phe Phe Glu Asp Glu Leu Arg Gly Ala Gly Phe
65 70 75 80
45 Gly Pro Glu His Asn Arg His Cys Leu Leu Thr Cys Glu Glu Cys Lys
85 90 95
Ile Lys His Gly Leu Ser Glu Lys Gly Asp Ser Glu Pro Ser Ala Ser
100 105 110

55 (2) INFORMATION FOR SEQ ID NO: 144:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

60

285

(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:
Met Lys Asn Asp Arg Asn Gln Gly Phe Ser Leu Leu Gln Leu Ile Asp
5 1 5 10 15
Trp Asn Lys Pro 20

10

(2) INFORMATION FOR SEQ ID NO: 145:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

Met Gly Thr Gln Pro Pro Val Val Ala Gly Phe Thr Ile Pro Met Leu
1 5 10 15
Gly Tyr Thr Val Arg Val Leu Thr Phe His Leu Ser Cys Ser
20 25 30

25

(2) INFORMATION FOR SEQ ID NO: 146:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

Met Lys Ile Pro Val Leu Pro Ala Val Val Leu Leu Ser Leu Leu Val
1 5 10 15

40

Leu His Ser Ala Gln Gly Ala Thr Leu Gly Gly Pro Glu Glu Ser
20 25 30

Thr Ile Glu Asn Tyr Ala Ser Arg Pro Glu Ala Phe Asn Thr Pro Phe
35 40 45

Leu Asn Ile Asp Lys Leu Arg Ser Ala Phe Lys Ala Asp Glu Phe Leu
50 55 60

Asn Trp His Ala Leu Phe Glu Ser Ile Lys Arg Lys Leu Pro Phe Leu
65 70 75 80

50

Asn Trp Asp Ala Phe Pro Lys Leu Lys Gly Leu Arg Ser Ala Thr Pro
85 90 95

55

Asp Ala Gln

(2) INFORMATION FOR SEQ ID NO: 147:

60

286

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

Met Val Trp Gly Leu Leu Gly
1 5

10

(2) INFORMATION FOR SEQ ID NO: 148:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

Met Leu Pro Leu Leu Ser Leu Leu Phe Leu Phe Ser Thr Val Ser
1 5 10 15

Ser Phe Cys Gly Met Pro Leu Arg Ala His Thr Arg Ala Xaa Ala His
20 25 30

25

Thr Arg Thr Phe Ala Ser Arg
35

(2) INFORMATION FOR SEQ ID NO: 149:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 131 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

Met Ile Cys Glu Thr Lys Ala Arg Lys Ser Ser Gly Gln Pro Gly Arg
1 5 10 15

Leu Pro Pro Thr Leu Ala Pro Gln Pro Pro Leu Pro Glu Thr
20 25 30

Ile Glu Arg Pro Val Gly Thr Gly Ala Met Val Ala Arg Ser Ser Asp
35 40 45

Leu Pro Tyr Leu Ile Val Gly Val Val Leu Gly Ser Ile Val Leu Ile
50 55 60

50

Ile Val Thr Phe Ile Pro Phe Cys Leu Trp Arg Ala Trp Ser Lys Gln
65 70 75 80

Lys His Thr Thr Asp Leu Gly Phe Pro Arg Ser Ala Leu Pro Pro Ser
85 90 95

Cys Pro Tyr Thr Met Val Pro Leu Gly Gly Leu Pro Gly His Gln Ala
100 105 110

Val Asp Ser Pro Thr Ser Val Ala Ser Val Asp Gly Pro Val Leu Met

287

115 120 125
Gly Ser Thr
130

(2) INFORMATION FOR SEQ ID NO: 150:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

15 Met Gly Ala Pro Ser Leu Thr Met Leu Leu Leu Lys Val Gln Pro
1 5 10 15
20 Arg Arg Thr Gln Ala Phe Asp Ala His Trp Val Gly Leu Pro Leu Leu
20 25 30

(2) INFORMATION FOR SEQ ID NO: 151:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

35 Met Cys Leu Ile Phe Leu Leu Leu Leu Ser Phe Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO: 152:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

45 His Pro His Gln Asp Ser Gln Pro
1 5

(2) INFORMATION FOR SEQ ID NO: 153:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

60

288

Met Asn Thr Ser Tyr Ile Leu Arg Leu Thr Val Val Ser Val Val
1 5 10 15
Ile Tyr Leu Ala Ile His Pro Leu Leu Ser Phe Ser Leu Gln Ser Pro
20 25 30

Leu Leu Val Pro Trp Arg Asp Cys Cys Gln Asn Ile Trp Lys Ser Gly
35 40 45

10 Ser Val Trp Tyr Lys Arg Trp Thr Leu Pro His Met Gln Val Cys Cys
50 55 60

Gln Asp Leu His
65

(2) INFORMATION FOR SEQ ID NO: 154:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

25 Met Leu Lys Ile Phe Gly Gln Trp Gln Asn Leu Asn Leu Ile Leu Thr
1 5 10 15

Ser Ile Arg Ile Leu Gln Arg Gln Asn Met
20 25

(2) INFORMATION FOR SEQ ID NO: 155:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 195 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

35 Met Asp Cys Gln Val Asn Asn Gly Ser Ser Leu Arg Asp Gln Cys Ile
1 5 10 15

45 Thr Asn Leu Leu Val Phe Gly Phe Leu Gln Ser Cys Ser Asp Asn Ser
20 25 30

Phe Arg Arg Gln Leu Asp Ala Leu Gly His Gln Leu Pro Val Leu Ala
35 40 45

50 Pro Gln Trp Gln Gly Tyr Asp Gln Leu Gln Thr Asp Gln Asn Arg Ser
50 55 60

Ser His Ser Arg Leu Gly Arg Ile Gln Ala Asp Ser Gln Ser Gln Gln
65 70 75 80

Asp Ile Ile Arg Asn Ile Ala Arg His Leu Ala Gln Val Gly Asp Ser
85 90 95

60 Met Asp Arg Ser Ile Pro Pro Gly Leu Val Asn Gly Leu Ala Leu Gln

289

100 105 110
 Leu Arg Asn Thr Ser Arg Ser Glu Glu Asp Arg Asn Arg Asp Leu Ala
 115 120 125

5 Thr Ala Leu Glu Gln Leu Leu Gln Ala Tyr Pro Arg Asp Met Glu Lys
 130 135 140

10 Glu Lys Thr Met Leu Val Leu Ala Leu Leu Ala Lys Lys Val Ala
 145 150 155 160

Ser His Thr Pro Ser Leu Leu Arg Asp Val Phe His Thr Thr Val Asn
 165 170 175

15 Phe Ile Asn Gln Asn Leu Arg Thr Tyr Val Arg Ser Leu Ala Arg Asn
 180 185 190

Gly Met Asp
 195

20

(2) INFORMATION FOR SEQ ID NO: 156:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 91 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

30 Met Ser Leu Ser Leu Val Ser Val Ser Val Gly Pro Ser Thr Leu Ala
 1 5 10 15

35 Cys Ser Phe Leu Arg Pro Lys Ala Arg Pro Ser Lys Arg Ser Pro Arg
 20 25 30

Asn Tyr Thr Asp Ser Thr Ser Pro Gly Gly Pro Arg Ala Pro Arg Gly
 35 40 45

40 Gly Ala Trp Arg Leu Ser Ser Gln Gln Asn Ser Ser Pro Lys Gly Val
 50 55 60

45 Ala Val Ala Lys Ala Ser Tyr Arg Pro Val Leu Cys Phe Leu Pro Gly
 65 70 75 80

Pro Trp Ser Ser Xaa Pro Xaa Ala Phe Leu Ile
 85 90

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

50 Met Gly Thr Leu Ser Ala Glu Cys Ser Gly Pro Ala Thr Leu Gly Leu
 1 5 10 15

WO 98/42738

290

Cys Leu Val Val Pro Trp Asn Ser Ser Gly Leu Ser Gln Pro Pro
 20 25 30

5

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 91 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

10 Met Lys Phe Leu Ala Val Leu Val Leu Leu Gly Val Ser Ile Phe Leu
 1 5 10 15

20 Val Ser Ala Gln Asn Pro Thr Thr Ala Ala Pro Ala Asp Thr Tyr Pro
 25 30

30 Ala Thr Gly Pro Ala Asp Asp Glu Ala Pro Asp Ala Glu Thr Thr Ala
 35 40 45

25 Ala Ala Thr Thr Ala Thr Thr Ala Ala Pro Thr Thr Ala Thr Thr Ala
 50 55 60

Ala Ser Thr Thr Ala Arg Lys Asp Ile Pro Val Leu Pro Lys Trp Val
 65 70 75 80

30 Gly Asp Leu Pro Asn Gly Arg Val Cys Pro Xaa
 85 90

(2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 89 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

45 Met Ile Ile Ser Leu Phe Ile Tyr Ile Phe Leu Thr Cys Ser Asn Thr
 1 5 10 15

Ser Pro Ser Tyr Gln Gly Thr Gln Leu Gly Leu Gly Leu Pro Ser Ala
 20 25 30

50 Gln Trp Trp Pro Leu Thr Thr Gly Arg Arg Met Gln Cys Cys Arg Leu Phe
 35 40 45

Cys Phe Leu Leu Gln Asn Cys Leu Phe Pro Phe Pro Leu His Leu Ile
 50 55 60

55 Gln His Asp Pro Cys Glu Leu Val Leu Thr Ile Ser Trp Asp Trp Ala
 65 70 75 80

Glu Ala Gly Ala Ser Leu Tyr Ser Pro
 85

60

291

(2) INFORMATION FOR SEQ ID NO: 160:

5 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 174 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

10 Met Ser Ser Ala Ala Asp His Trp Ala Trp Leu Leu Val Leu Ser
1 5 10 1515 Phe Val Phe Gly Cys Asn Val Leu Arg Ile Leu Leu Pro Ser Phe Ser
20 25 30Ser Phe Met Ser Arg Val Leu Gln Lys Asp Ala Gln Gln Ser Gln
35 40 4520 Met Arg Ala Gln Ile Gln Asp Met Lys Met Lys Gln Leu Ser Thr Val Asn
50 55 6025 Met Met Asp Gln Phe Ala Arg Tyr Ala Arg Leu Gln Arg Lys Ile Asn
65 70 75 80Lys Met Thr Asp Lys Leu Lys Thr His Val Lys Ala Arg Thr Ala Gln
85 90 9530 Leu Ala Lys Ile Lys Trp Val Ile Ser Val Ala Phe Tyr Val Leu Gln
100 105 110Ala Ala Leu Met Ile Ser Leu Ile Trp Lys Tyr Tyr Ser Val Pro Val
115 120 12535 Ala Val Val Pro Ser Lys Trp Ile Thr Pro Leu Asp Arg Leu Val Ala
130 135 140Phe Pro Thr Arg Val Ala Gly Gly Val Gly Ile Thr Cys Trp Ile Leu
145 150 155 16040 Val Cys Asn Lys Val Val Ala Ile Val Leu His Pro Phe Ser
165 170

45 (2) INFORMATION FOR SEQ ID NO: 161:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

55 Met Gly Lys Leu Ile Asn Ile Val Ile Arg Lys Pro Leu Leu Leu Leu
1 5 10 15Leu Val Gln Cys Gln Asn Cys Cys Arg Lys Asn Met Leu Tyr Asn Ile
20 25 30

60 Phe Leu Asn Ile His Asn Ile His Lys Phe Ser Asn His

292

35 40 45

5 (2) INFORMATION FOR SEQ ID NO: 162:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

10 Met Val Ala Ser Thr Leu Val Thr Asn Leu Phe Gly Val Ala Phe Ala
1 5 10 1515 Thr Thr Ala Ala Thr Arg Ala
20

20 (2) INFORMATION FOR SEQ ID NO: 163:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

30 Met Leu Met Ala Pro Val Val Cys Leu Ser Phe Ser Pro Cys Pro Ala
1 5 10 15Asp Thr Ser Leu Thr Gly Asp Gly Leu Lys Ala Gly Leu Gln Arg Gly
20 25 3035 Xaa Ala Leu Val Thr Leu Phe Asp Ser Val Thr His Phe Leu Ala His
35 40 45Thr Leu Phe Gln Leu Leu Asp Phe Gln Leu Ala Phe Leu Arg Ser Gly
50 55 6040 Lys Gln Thr Ala Pro His
65 70

45 (2) INFORMATION FOR SEQ ID NO: 164:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 323 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

55 Met Leu Leu Leu Leu Leu Leu Leu Gly Ser Gly Gln Gly Pro Gln Gln
1 5 10 15Val Gly Ala Gly Gln Thr Phe Gln Tyr Leu Lys Arg Gln His Ser Leu
20 25 30

60 Ser Lys Pro Tyr Gln Gly Val Gly Thr Gly Ser Ser Ser Leu Trp Asn

293

294

35 40 45
 Leu Met Gly Asn Ala Met Val Met Thr Gln Tyr Ile Arg Leu Thr Pro
 50 55 60
 5 Asp Met Gln Ser Lys Gln Gly Ala Leu Trp Asn Arg Val Pro Cys Phe
 65 70 75 80
 10 Leu Arg Asp Trp Glu Leu Gln Val His Phe Lys Ile His Gly Gln Gly
 85 90 95
 Lys Lys Asn Leu His Gly Asp Gly Leu Ala Ile Trp Tyr Thr Arg Asn
 100 105 110
 15 Arg Met Gln Pro Gly Pro Val Phe Gly Asn Met Asp Lys Phe Val Gly
 115 120 125
 Leu Gly Val Phe Val Asp Thr Tyr Pro Asn Glu Glu Lys Gln Gln Glu
 130 135 140
 20 Arg Val Phe Pro Tyr Ile Ser Ala Met Val Asn Asn Gly Ser Leu Ser
 145 150 155 160
 Tyr Asp His Glu Arg Asp Gly Arg Pro Thr Glu Leu Gly Gly Cys Thr
 165 170 175
 25 Ala Ile Val Arg Asn Leu His Tyr Asp Thr Phe Leu Val Ile Arg Tyr
 180 185 190
 30 Val Lys Arg His Leu Thr Ile Met Met Asp Ile Asp Gly Lys His Glu
 195 200 205
 Trp Arg Asp Cys Ile Glu Val Pro Gly Val Arg Leu Pro Arg Gly Tyr
 210 215 220
 35 Tyr Phe Gly Thr Ser Ser Ile Thr Gly Asp Leu Ser Asp Asn His Asp
 225 230 235 240
 Val Ile Ser Leu Lys Leu Phe Glu Leu Thr Val Glu Arg Thr Pro Glu
 245 250 255
 Glu Glu Lys Leu His Arg Asp Val Phe Leu Pro Ser Val Asp Asn Met
 260 265 270
 45 Lys Leu Pro Glu Met Thr Ala Pro Leu Pro Pro Leu Ser Gly Leu Ala
 275 280 285
 Leu Phe Leu Ile Val Phe Phe Ser Leu Val Phe Ser Val Phe Ala Ile
 290 295 300
 50 Val Ile Gly Ile Ile Leu Tyr Asn Lys Trp Gln Glu Gln Ser Arg Lys
 305 310 315 320
 Arg Phe Tyr

(2) INFORMATION FOR SEQ ID NO: 165:

60

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 321 amino acids

(B) TYPE: amino acid

(C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

5 Met Pro Ser Glu Tyr Thr Tyr Val Lys Leu Arg Ser Asp Cys Ser Arg
 1 5 10 15
 10 Pro Ser Leu Gln Trp Tyr Thr Arg Ala Gln Ser Lys Met Arg Arg Pro
 20 25 30
 Ser Leu Leu Leu Lys Asp Ile Leu Lys Cys Thr Leu Leu Val Phe Gly
 35 40 45
 15 Val Trp Ile Leu Tyr Ile Leu Lys Leu Asn Tyr Thr Thr Glu Cys
 50 55 60
 Asp Met Lys Lys Met His Tyr Val Asp Pro Asp His Val Lys Arg Ala
 65 70 75 80
 Gln Lys Tyr Ala Gln Gln Val Leu Gln Lys Glu Cys Arg Pro Lys Phe
 85 90 95
 25 Ala Lys Thr Ser Met Ala Leu Leu Phe Glu His Arg Tyr Ser Val Asp
 100 105 110
 Leu Leu Pro Phe Val Gln Lys Xaa Pro Lys Asp Ser Glu Ala Glu Ser
 115 120 125
 30 Lys Tyr Asp Pro Pro Phe Gly Phe Arg Lys Phe Ser Ser Lys Val Gln
 130 135 140
 Thr Leu Leu Glu Leu Leu Pro Glu His Asp Leu Pro Glu His Leu Lys
 145 150 155 160
 Ala Lys Thr Cys Arg Arg Cys Val Val Ile Gly Ser Gly Ile Leu
 165 170 175
 40 His Gly Leu Glu Leu Gly His Thr Leu Asn Gln Phe Asp Val Val Ile
 180 185 190
 Arg Leu Asn Ser Ala Pro Val Glu Gly Tyr Ser Glu His Val Gly Asn
 195 200 205
 45 Lys Thr Thr Ile Arg Met Thr Tyr Pro Glu Gly Ala Pro Leu Ser Asp
 210 215 220
 Leu Glu Tyr Tyr Ser Asn Asp Leu Phe Val Ala Val Leu Phe Lys Ser
 225 230 235 240
 Val Asp Phe Asn Trp Leu Gln Ala Met Val Lys Lys Glu Thr Leu Pro
 245 250 255
 55 Phe Trp Val Arg Leu Phe Phe Trp Lys Gln Val Ala Glu Lys Ile Pro
 260 265 270
 Leu Gln Pro Lys His Phe Arg Ile Leu Asn Pro Val Ile Ile Lys Glu
 275 280 285

295

Thr Ala Phe Xaa His Pro Ser Val Leu Arg Ala Ser Val Lys Val Leu
290 295 300

Gly Ala Glu Ile Arg Thr Ser Pro Gln Ser Val Ser Leu Pro Leu Ser
305 310 315 320

Xaa

10

(2) INFORMATION FOR SEQ ID NO: 166:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

Met Thr Leu Asp Val Gln Thr Val Val Val Phe Ala Val Ile Val Val
1 5 10 15

Leu Leu Leu Val Asn Val Ile Leu Met Phe Phe Leu Gly Thr Arg
20 25 30

25

(2) INFORMATION FOR SEQ ID NO: 167:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

Met Leu Pro Leu Leu Phe Cys Ala Phe Cys Leu His Lys Leu Gly Pro
1 5 10 15

Leu Leu Phe Leu Tyr Asp Val Leu Met Xaa His Glu Ala Val Met Arg
20 25 30

Thr His Gln Ile Gln Leu Pro Asp Pro Glu Phe Pro Ser Gln Gln Asn
35 40 45

Gln Val Leu Asn Lys Thr Leu Phe Asn Lys Leu Lys Lys Lys Lys
50 55 60

Lys Lys Lys Xaa Xaa Lys Lys
65 70

50

(2) INFORMATION FOR SEQ ID NO: 168:

55

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 282 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

60

296

Met Ala Ser Arg Gly Arg Arg Pro Glu His Gly Gly Pro Pro Glu Leu
1 5 10 15

Phe Tyr Asp Glu Thr Glu Ala Arg Lys Tyr Val Arg Asn Ser Arg Met
20 25 30

Ile Asp Ile Gln Thr Arg Met Ala Gly Arg Ala Leu Glu Leu Leu Tyr
35 40 45

Leu Pro Glu Asn Lys Pro Cys Tyr Leu Leu Asp Ile Gly Cys Gly Thr
50 55 60

Gly Leu Ser Gly Ser Tyr Leu Ser Asp Glu Gly His Tyr Trp Val Gly
65 70 75 80

Leu Asp Ile Ser Pro Ala Met Leu Asp Glu Ala Val Asp Arg Gly Ile
85 90 95

Glu Gly Asp Leu Leu Gly Asp Met Gly Gln Gly Ile Pro Phe Lys
100 105 110

Pro Gly Thr Phe Asp Gly Cys Ile Ser Ile Ser Ala Val Gln Trp Leu
115 120 125

Cys Asn Ala Asn Lys Lys Ser Glu Asn Pro Ala Lys Arg Leu Tyr Cys
130 135 140

Phe Phe Ala Ser Leu Phe Ser Val Leu Val Arg Gly Ser Arg Ala Val
145 150 155 160

Leu Gln Leu Tyr Pro Glu Asn Ser Glu Gln Leu Leu Ile Thr Thr
165 170 175

Gln Ala Thr Lys Ala Gly Phe Ser Gly Gly Met Val Val Asp Tyr Pro
180 185 190

Asn Ser Ala Lys Ala Lys Lys Phe Tyr Leu Cys Leu Phe Ser Gly Pro
195 200 205

Ser Thr Phe Ile Pro Glu Gly Leu Ser Glu Asn Gln Asp Glu Val Glu
210 215 220

Pro Arg Glu Ser Val Phe Thr Asn Glu Arg Phe Pro Leu Arg Met Ser
225 230 235 240

Arg Arg Gly Met Val Arg Lys Ser Arg Ala Trp Val Leu Glu Lys Lys
245 250 255

Glu Arg His Arg Arg Gln Gly Arg Glu Val Arg Pro Asp Thr Gln Tyr
260 265 270

Thr Gly Arg Lys Arg Lys Pro Arg Phe Xaa
275 280

55

(2) INFORMATION FOR SEQ ID NO: 169:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

60

297

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

5 Met Leu Gly Lys Thr Lys Phe Gln Ser Tyr Lys Ser Phe Ser Arg Lys 15
1
Leu Met Val Cys Pro Ser Thr 20

10

(2) INFORMATION FOR SEQ ID NO: 170:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 328 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

20 Met Trp Arg Pro Ser Val Leu Leu Leu Leu Leu Leu Leu Leu Arg His Gly 15
1
Ala Gln Gly Lys Pro Ser Pro Asp Ala Gly Pro His Gly Gln Gly Arg 25
25 Val His Gln Ala Ala Pro Leu Ser Asp Ala Pro His Asp Asp Ala His 35
30 Gly Asn Phe Gln Tyr Asp His Glu Ala Phe Leu Gly Arg Glu Val Ala 50
55
Lys Glu Phe Asp Gln Leu Thr Pro Glu Glu Ser Gln Ala Arg Leu Gly 70
65
Arg Ile Val Asp Arg Met Asp Arg Ala Gly Asp Gly Asp Gly Trp Val 85
90
Ser Leu Ala Glu Leu Arg Ala Trp Ile Ala His Thr Gln Gln Arg His 100
105
Ile Arg Asp Ser Val Ser Ala Ala Trp Asp Thr Tyr Asp Thr Asp Arg 110
115
Asp Gly Arg Val Gly Trp Glu Glu Leu Arg Arg Asn Ala Thr Tyr Gly His 120
130
Tyr Ala Pro Gly Glu Glu Phe His Asp Val Glu Asp Ala Glu Thr Tyr 135
140
Lys Lys Met Leu Ala Arg Asp Glu Arg Arg Phe Arg Val Ala Asp Gln 145
150
Asp Gly Asp Ser Met Ala Thr Arg Glu Glu Leu Thr Ala Phe Leu His 155
160
Pro Glu Glu Phe Pro His Met Arg Asp Ile Val Ile Ala Glu Thr Leu 165
170
Glu Asp Leu Asp Arg Asn Lys Asp Gly Tyr Val Gln Val Glu Glu Tyr 175
180
185
190
195
200
205

298

210 215 220
Ile Ala Asp Leu Tyr Ser Ala Glu Pro Gly Glu Glu Pro Ala Trp 225
230
Val Gln Thr Glu Arg Gln Gln Phe Arg Asp Phe Arg Asp Leu Asn Lys 235
240
245
Asp Gly His Leu Asp Gly Ser Glu Val Gly His Trp Val Leu Pro Pro 250
255
260
Ala Gln Asp Gln Pro Leu Val Glu Ala Asn His Leu Leu His Glu Ser 265
270
275
Asp Thr Asp Lys Asp Gly Arg Leu Ser Lys Ala Xaa Ile Leu Gly Asn 280
285
Trp Asn Met Phe Val Gly Ser Gln Ala Thr Asn Tyr Gly Glu Asp Leu 290
300
305
Thr Arg His His Asp Glu Leu Xaa 310
315
320
325

(2) INFORMATION FOR SEQ ID NO: 171:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

25 Met Cys Trp Leu Arg Ala Trp Xaa Gln Ile Xaa Leu Pro Val Phe Xaa 1
5
10
15
Ser Xaa Phe Leu Ile Gln Leu Leu Ile Ser Phe Ser Glu Asn Gly Phe 20
25
30
Ile His Ser Pro Arg Asn Asn Gln Lys Pro Arg Asp Gly Asn Xaa Glu 35
40
45
Glu Cys Ala Val Lys Lys Ser Cys Gln Leu Cys Thr Glu Asp Lys Lys 50
55
Tyr Met Met Asn Arg 60
65

(2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

50 Met Trp Leu Phe Ile Leu Leu Ser Leu Ala Leu Ile Ser Asp Ala Met 1
5
10
15

299

Val Met Asp Glu Lys Val Lys Arg Ser Phe Val Leu Asp Thr Ala Ser
20 25 30

5 Ala Ile Cys Asn Tyr Asn Ala His Tyr Lys Asn His Pro Lys Tyr Trp
35 40 45

10 Cys Arg Gly Tyr Phe Arg Asp Tyr Cys Asn Ile Ile Ala Phe Ser Pro
50 55 60

Asn Ser Thr Asn His Val Ala Leu Lys Asp Thr Gly Asn Gln Leu Ile
65 70 75 80

15 Val Thr Met Ser Cys Leu Asn Lys Glu Asp Thr Gly Trp Tyr Trp Cys
85 90 95

Gly Ile Gln Arg Asp Phe Ala Arg Asp Asp Met Asp Phe Thr Glu Leu
100 105 110

20 Ile Val Thr Asp Asp Lys Gly Thr Trp Pro Met Thr Leu Val Trp Glu
115 120 125

Arg Leu Ser Gly Thr Lys Pro Glu Ala Ala Arg Leu Pro Lys Leu Ser
130 135 140

25 Ala Arg Leu Thr Ala Pro Gly Arg Pro Phe Ser Ser Phe Ala Tyr Xaa
145 150 155 160

30

(2) INFORMATION FOR SEQ ID NO: 173:

35

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 123 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

40 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

Met Ala Xaa His Phe Leu Leu Val Ala Leu Gln Ser Val Pro His Cys
1 5 10 15

45 Pro His Leu Leu Glu Glu Glu His Lys Leu Cys Lys Val Ser His Phe
20 25 30

Ser Gly Val Thr Leu Val Thr Ser Arg Gln Asp Ser Ser Tyr Val
35 40 45

50 Pro Val Gln Thr Leu Phe Ile His Leu Gly Pro Trp Ala Trp Asp Leu
50 55 60

55 Xaa Pro Cys Thr Ala Glu Asp Pro Glu Ala Glu Arg Ser Leu Arg Leu
65 70 75 80

Cys His Ser His Leu Ala Arg Xaa Asn Val Ser Pro Ser Gln Ala Ala
85 90 95

60 Glu Gly Xaa Xaa Xaa Arg Gly Cys Gln His Arg Gly Ser Arg Glu Leu

300

100 105 110
Thr Phe Leu Ser Ala Glu Asn Glu Ala Gly Ile
115 120

5

(2) INFORMATION FOR SEQ ID NO: 174:

10

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 129 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

15 Met Lys Val Gly Ala Arg Ile Arg Val Lys Met Ser Val Asn Lys Ala
1 5 10 15

20 His Pro Val Val Ser Thr His Trp Arg Trp Pro Ala Glu Trp Pro Gln
20 25 30

Met Phe Leu His Leu Ala Gln Glu Pro Arg Thr Glu Val Lys Ser Arg
35 40 45

25 Pro Leu Gly Leu Ala Gly Phe Ile Arg Gln Asp Ser Lys Thr Arg Lys
50 55 60

Pro Leu Glu Gln Glu Thr Ile Met Ser Ala Ala Asp Thr Ala Leu Trp
65 70 75 80

30 Pro Tyr Gly His Gly Asn Arg Arg Glu His Gln Glu Asn Glu Leu Lys
85 90 95

35 Tyr Leu Gln Tyr Lys Asp Met His Leu Leu Asp Ser Gly Gln Ser Leu
100 105 110

Gly His Thr His Thr Leu Gln Gly Ser His Asn Leu Thr Ala Leu Asn
115 120 125

40 Ile

45 (2) INFORMATION FOR SEQ ID NO: 175:

50

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 372 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

Met Ala Tyr His Ser Phe Leu Val Glu Pro Ile Ser Cys His Ala Trp
1 5 10 15

55 Asn Lys Asp Arg Thr Thr Gln Ile Ala Ile Cys Pro Asn Asn His Glu Val
20 25 30

His Ile Tyr Glu Lys Ser Gly Ala Lys Trp Thr Thr Lys Val His Glu Leu
35 40 45

301

302

Lys Glu His Asn Gly Gln Val Thr Gly Ile Asp Trp Trp Ala Pro Glu Ser
 50 55 60
 5 Asn Arg Ile Val Thr Cys Gly Thr Asp Arg Asn Ala Tyr Val Thr Thr
 65 70 75 80
 Leu Lys Gly Arg Thr Trp Lys Pro Thr Leu Val Ile Leu Arg Ile Asn
 85 90 95
 10 Arg Ala Ala Arg Cys Val Arg Trp Ala Pro Asn Glu Asn Lys Phe Ala
 100 105 110
 Val Gly Ser Gly Ser Arg Val Ile Ser Ile Cys Tyr Phe Glu Gln Glu
 115 120 125
 15 Asn Asp Trp Trp Val Cys Lys His Ile Lys Lys Pro Ile Arg Ser Thr
 130 135 140
 20 Val Leu Ser Leu Asp Trp His Pro Asn Asn Val Leu Leu Ala Ala Gly
 145 150 155 160
 Ser Cys Asp Phe Lys Cys Arg Ile Phe Ser Ala Tyr Ile Lys Glu Val
 165 170 175
 25 Glu Glu Arg Pro Ala Pro Thr Pro Trp Gly Ser Lys Met Pro Phe Gly
 180 185 190
 Glu Leu Met Phe Glu Ser Ser Ser Cys Gly Trp Val His Gly Val
 195 200 205
 30 Cys Phe Ser Ala Ser Gly Ser Arg Val Ala Trp Val Ser His Asp Ser
 210 215 220
 35 Thr Val Cys Leu Ala Asp Ala Asp Lys Lys Met Ala Val Ala Thr Leu
 225 230 235 240
 Ala Ser Glu Thr Leu Pro Leu Leu Ala Leu Thr Phe Ile Thr Asp Asn
 245 250 255
 40 Ser Leu Val Ala Ala Gly His Asp Cys Phe Pro Val Leu Phe Thr Tyr
 260 265 270
 45 Asp Ala Ala Gly Met Leu Ser Phe Gly Arg Leu Asp Val Pro
 275 280 285
 Lys Gln Ser Ser Gln Arg Gly Leu Thr Ala Arg Glu Arg Phe Gln Asn
 290 295 300
 50 Leu Asp Lys Lys Ala Ser Ser Glu Gly Thr Ala Ala Gly Ala Gly
 305 310 315 320
 Leu Asp Ser Leu His Lys Asn Ser Val Ser Gln Ile Ser Val Leu Ser
 325 330 335
 55 Gly Gly Lys Ala Lys Cys Ser Gln Phe Cys Thr Thr Gly Met Asp Gly
 340 345 350
 Gly Met Ser Ile Trp Asp Val Lys Ser Leu Glu Ser Ala Leu Lys Asp
 355 360 365

Leu Lys Ile Lys
 370
 5
 (2) INFORMATION FOR SEQ ID NO: 176:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 216 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 176:
 1 Met Trp Ser Ile Gly Ala Gly Ala Leu Gly Ala Ala Leu Ala Leu
 5 10 15
 Leu Leu Ala Asn Thr Asp Val Phe Leu Ser Lys Pro Gln Lys Ala Ala
 20 25 30
 Leu Glu Tyr Leu Glu Asp Ile Asp Leu Lys Thr Leu Glu Lys Glu Pro
 35 40 45
 Arg Thr Phe Lys Ala Lys Glu Leu Trp Glu Lys Asn Gly Ala Val Ile
 50 55 60
 Met Ala Val Arg Arg Pro Gly Cys Phe Leu Cys Arg Glu Glu Ala Ala
 65 70 75 80
 Asp Leu Ser Ser Leu Lys Ser Met Leu Asp Gln Leu Gly Val Pro Leu
 85 90 95
 Tyr Ala Val Val Lys Glu His Ile Arg Thr Glu Val Lys Asp Phe Gln
 100 105 110
 Pro Tyr Phe Lys Gly Glu Ile Phe Leu Asp Glu Lys Lys Phe Tyr
 115 120 125
 Gly Pro Gln Arg Arg Lys Met Met Phe Met Gly Phe Ile Arg Leu Gly
 130 135 140
 Val Trp Tyr Asn Phe Phe Arg Ala Trp Asn Gly Gly Phe Ser Gly Asn
 145 150 155 160
 Leu Glu Gly Glu Gly Phe Ile Leu Glu Gly Val Phe Val Val Gly Ser
 165 170 175
 Gly Lys Gln Gly Ile Leu Leu Glu His Arg Glu Lys Glu Phe Gly Asp
 180 185 190
 Lys Val Asn Leu Ser Val Leu Glu Ala Ala Lys Met Ile Lys Pro
 195 200 205
 Gln Thr Leu Ala Ser Glu Lys Lys
 210 215
 55
 (2) INFORMATION FOR SEQ ID NO: 177:
 60

303

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 55 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

5 Met Lys Pro Val Ser Arg Arg Thr Leu Asp Trp Ile Tyr Ser Val Leu
1 5 10 15

10 Leu Leu Ala Ile Val Leu Ile Ser Trp Gly Cys Ile Ile Tyr Ala Ser
20 25 30

15 Met Val Ser Ala Arg Arg Gln Leu Arg Lys Tyr Pro Asp Lys Ile
35 40 45

Phe Gly Thr Asn Gln Asn Leu
50 55

(2) INFORMATION FOR SEQ ID NO: 178:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

25 Met Ala Ala Asn Thr Phe Val Leu Ile Met Gly Ile Pro Thr Ser Ala
1 5 10 15

30 Asn Ala Xaa Arg Asp Leu Phe
20

(2) INFORMATION FOR SEQ ID NO: 179:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

35 Met Ser Ile Cys His Arg Gly Thr Gly Ile Ala Leu Ser Ala Gly Val
1 5 10 15

40 Ser Leu Phe Gly Met Ser Ala Leu Leu Leu Pro Gly Asn Phe Gln Ser
20 25 30

45 Tyr Leu Gln Leu Val Lys Ser Leu Cys Leu Gly Pro Ala Leu Ile His
35 40 45

50 Thr Ala Lys Phe Ala Leu Val Phe Pro Leu Met Tyr His Thr Trp Asn
50 55 60

55 Gly Ile Arg His Leu Met Trp Asp Leu Gly Lys Leu Lys Ile Pro
65 70 75 80

60 Gln Leu Tyr Gln Ser Gly Val Val Leu Val Leu Thr Val Leu Ser

304

85 90 95

Ser Met Gly Leu Ala Ala Met
100

(2) INFORMATION FOR SEQ ID NO: 180:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

10 Met Thr Lys Ala Ser Ser Leu Trp Pro Leu Lys Thr Thr Cys Gln Ile
1 5 10 15

15 Ser Gly Thr Val Phe Phe Leu Phe Phe Ser Cys Phe Leu Met
20 25 30

20 Gln Ala Gln Cys Asp Lys Phe Val Gly Trp Asp Phe Phe Phe Leu
35 40 45

25

(2) INFORMATION FOR SEQ ID NO: 181:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 96 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

30 Met Arg Arg Ala Leu Ile Pro Pro Cys Arg Gly Gly Pro Ser Ala Ser
1 5 10 15

35 Asp Xaa Cys Cys Ser Cys Ser Pro Ser Gly Phe Ser Ala Gly Arg Gly
20 25 30

40 Arg Cys Pro Val Gln Gly Cys Leu Arg Pro His Arg Val Gln Leu Leu
35 40 45

45 Arg Arg Trp Gly Pro Gly Ser Pro Ala Gly Gln Arg Leu Ser Lys Gly
50 55 60

50 Phe Gln Leu Leu Arg Trp Trp Gly Ser Pro Ala Pro Gln Pro
65 70 75 80

55 Arg Lys Gly Pro Phe Pro Pro Asp Pro Pro Trp Pro Val Thr Leu
85 90 95

60

(2) INFORMATION FOR SEQ ID NO: 182:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

Met Leu Glu Thr Thr Lys His Val Gln Ile Ala Cys Met Leu Leu Leu 1
1 5 10 15

Thr Cys Gln Ile Phe Leu Pro Ser Ser Leu Ser Pro Ser Phe Ile His 20
25 30

Ser Leu Thr Asp Ser Phe Ile Pro Leu Lys Lys Leu Tyr Val Cys Phe 35
40 45

Val Gln Ser Thr Leu Leu Lys Ala Ala Gly Tyr Lys Ser Ile Ser Glu 50
55 60

Ala Leu Gly Phe Asp Xaa Leu Leu Cys Ser Ser Ala Arg Phe Val Trp 65
70 75 80

Ile Cys His Thr Tyr Ser Arg Pro Leu Val Thr Cys Ala Leu His 85
90 95

(2) INFORMATION FOR SEQ ID NO: 183:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

Met Ser Val Ile Gly Gly Leu Leu Leu Val Val Ala Leu Gly Pro Gly 1
5 10 15

Gly Val Ser Met Asp Glu Lys Lys Lys Glu Trp 20
25

(2) INFORMATION FOR SEQ ID NO: 184:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

Met Ser Gly Gly Leu Ser Phe Leu Leu Val 1
5 10

(2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 65 amino acids

- (B) TYPE: amino acid

- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

Met Phe Ala Asp Phe Ile Val Val Thr Ala Thr Val Gln Arg Cys Pro 1
5 10 15

Gly Ser Pro Pro Leu Ser Glu Ile Leu Trp Lys Asp Glu Pro Phe Ala 20
25 30

Ile Ser Ser His Ala Gly Leu Pro Trp Leu Ser Ser Trp Pro Ala Pro 35
40 45

Pro Trp Thr Trp Ser Trp Ile Ser Arg Arg Glu His Gly Arg Gly 50
55 60

Ser 65

(2) INFORMATION FOR SEQ ID NO: 186:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids

- (B) TYPE: amino acid

- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

Met Val Glu Ser Val Met Pro Val Val Val Cys Thr Leu Ser Pro Gly 1
5 10 15

Ile Asp Ser Ser Pro Ser 20

(2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids

- (B) TYPE: amino acid

- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

Met Asp Val Leu Phe Val Ala Ile Phe Ala Val Pro Leu Ile Leu Gly 1
5 10 15

Gln Glu Tyr Glu Asp Glu Arg Leu Gly Glu Asp Glu Tyr Tyr Gln 20
25 30

Val Val Tyr Tyr Thr Val Thr Pro Ser Tyr Asp Phe Ser Ala 35
40 45

Asp Phe Thr Ile Asp Tyr Ser Ile Phe Glu Ser Glu Asp Arg Leu Asn 50
55 60

Arg Leu Asp Lys Asp Ile Thr Glu Ala Ile Glu Thr Thr Ile Ser Leu 65
70 75 80

10	Met Tyr Phe Met	130
5	Thr Thr Glu Pro Gln Ser Pro Asp Leu Asn Asp Ala Val Ser Ser Leu	110
	100	105
	Arg Ser Pro Ile Pro Leu Leu Leu Ser Cys Ala Phe Val Gln Val Gly	125
	115	120
	Glu Thr Ala Arg Ala Asp His Pro Lys Pro Val Thr Val Lys Pro Val	95
	85	90

15 (2) INFORMATION FOR SEQ ID NO: 188:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 69 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

[illegible]

40

(2) INFORMATION FOR SEQ ID NO: 189:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

50 Met Ser Leu Leu Ser Pro Ala Ile Pro Ala Leu Thr Leu Ile Phe Ile
1 5 10 15

Leu Met Phe Phe Ser Phe Pro Phe Arg Ala His Thr Val Thr Ile
20 25 30

55 Val Ala Ser Gly Phe Leu Gly Leu Ser Pro Leu Cys Gly
35 40 45

60 (2) INFORMATION FOR SEQ ID NO: 190:

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 65 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 190

[illegible]

25 (2) INFORMATION FOR SEQ ID NO: 191:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

30 Met Asn Leu Leu Gly Met Ile Phe Ser Met Cys Gly Leu Met Leu Lys
1 5 10 15

35 Leu Lys Trp Cys Ala Trp Val Ala Val Tyr Cys Ser Phe Ile Ser Phe
20 25 30

40 Ala Asn Ser Arg Ser Ser Glu Asp Thr Lys Gln Met Met Ser Ser Phe
35 40 45

Met Xaa
50

45

(2) INFORMATION FOR SEQ ID NO: 192:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 170 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

Met	Leu	Leu	Asn	Val	Ala	Leu	Val	Ala	Leu	Val	Leu	Gly	Ala	Tyr
1					5				10				15	
Arg	Leu	Tyr	Val	Arg	Tyr	Gly	Arg	Arg	Gly	Leu	Gly	Ala	Gly	Ala
20								25					30	

60

309

Ala Gly Glu Glu Ser Pro Ala Thr Ser Leu Pro Arg Met Lys Lys Arg
35 40 45

Asp Phe Ser Leu Glu Gln Leu Arg Gln Tyr Asp Gly Ser Arg Asn Pro
5 50 55 60

Arg Ile Leu Leu Ala Val Asn Gly Lys Val Phe Asp Val Thr Lys Gly
65 70 75 80

Ser Lys Phe Tyr Gly Pro Ala Gly Pro Tyr Gly Ile Phe Ala Gly Arg
10 85 90 95

Asp Ala Ser Arg Gly Leu Ala Thr Phe Cys Leu Asp Lys Asp Ala Leu
100 105 110

Arg Asp Glu Tyr Asp Asp Leu Ser Asp Leu Asn Ala Val Gln Met Glu
115 120 125

Ser Val Arg Glu Trp Glu Met Gln Phe Lys Glu Lys Tyr Asp Tyr Val
130 135 140

Gly Arg Leu Leu Lys Pro Gly Glu Glu Pro Ser Glu Tyr Thr Asp Glu
145 150 155 160

Glu Asp Thr Lys Asp His Asn Lys Gln Asp
165 170

(2) INFORMATION FOR SEQ ID NO: 193:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

Met Thr Tyr Phe Ser Gly Leu Leu Val Ile Leu Ala Phe Ala Ala Trp
1 5 10 15

Val Ala Leu Ala Glu Gly Leu Gly Val Ala Val Tyr Ala Ala Val
20 25 30

Leu Leu Gly Ala Gly Cys Ala Thr Ile Leu Val Thr Ser Leu Ala Met
35 40 45

Thr Ala Asp Leu Ile Gly Pro His Thr Asn Ser Gly Leu Ser Cys Thr
50 55 60

Ala Pro
65

(2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 92 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

310

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

Met Ala Ala Gly Pro Ser Gly Cys Leu Val Pro Ala Phe Gly Leu Arg
1 5 10 15

Leu Leu Leu Ala Thr Val Leu Gln Ala Val Ser Ala Phe Gly Ala Glu
20 25 30

Phe Ser Ser Glu Ala Cys Arg Glu Leu Gly Phe Ser Ser Asn Leu Leu
35 40 45

Cys Ser Ser Cys Asp Leu Leu Gly Gln Phe Asn Leu Leu Gln Leu Asp
50 55 60

Pro Asp Cys Arg Gly Cys Cys Gln Glu Glu Ala Gln Phe Glu Thr Lys
65 70 75 80

Lys Leu Tyr Ala Gly Ala Ile Leu Glu Val Cys Gly
85 90

(2) INFORMATION FOR SEQ ID NO: 195:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 176 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

Met Arg Gly Ser His Leu Arg Leu Leu Pro Tyr Leu Val Ala Ala Asn
1 5 10 15

Pro Val Asn Tyr Gly Arg Pro Tyr Arg Leu Ser Cys Val Glu Ala Phe
20 25 30

Ala Ala Thr Phe Cys Ile Val Gly Phe Pro Asp Leu Ala Val Ile Leu
35 40 45

Leu Arg Lys Phe Lys Trp Gly Lys Gly Phe Leu Asp Leu Asn Arg Gln
50 55 60

Leu Leu Asp Lys Tyr Ala Ala Cys Gly Ser Pro Glu Glu Val Leu Gln
65 70 75 80

Ala Glu Gln Glu Phe Leu Ala Asn Ala Lys Glu Ser Pro Gln Glu Glu
85 90 95

Glu Ile Asp Pro Phe Asp Val Asp Ser Gly Arg Glu Phe Gly Asn Pro
100 105 110

Asn Arg Pro Val Ala Ser Thr Arg Leu Pro Ser Asp Thr Asp Ser Ser
115 120 125

Asp Ala Ser Glu Asp Pro Gly Pro Xaa Ala Glu Arg Gly Gly Ala Ser
130 135 140

Ser Ser Cys Cys Glu Glu Gln Thr Gln Gly Arg Gly Ala Glu Ala
145 150 155 160

311

Arg Ala Pro Ala Glu Val Trp Lys Gly Ile Lys Lys Arg Gln Arg Asp
165 170 175

5

(2) INFORMATION FOR SEQ ID NO: 196:

10

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

Met Ser Asn Ala Cys Lys Glu Leu Ala Ile Phe Leu Thr Thr Gly Ile
1 5 10 15

Val Val Ser Ala Phe Gly Leu Pro Ile Val Phe Ala Arg Ala His Leu
20 25 30

Ile Glu Trp Gly Ala Cys Ala Leu Val Leu Thr Gly Asn Thr Val Ile
35 40 45

Phe Ala Thr Ile Leu Gly Phe Phe Leu Val Phe Gly Ser Asn Asp Asp
50 55 60

Phe Ser Trp Gln Gln Trp
65 70

(2) INFORMATION FOR SEQ ID NO: 197:

35

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

Met Thr Leu Leu Ile Ile Phe Leu Pro Phe Xaa Phe Thr Thr Xaa Thr
1 5 10 15

Asn Ser Gly Gly Ser Phe Pro Val Arg
20 25

(2) INFORMATION FOR SEQ ID NO: 198:

50

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 73 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

Met Lys Gly Glu Leu Leu Pro Phe Leu Thr Thr Val Trp Leu Trp
1 5 10 15

312

Leu Tyr Lys Leu Xaa Phe Gly Glu Ser Pro Arg Tyr Pro Asn Val Ile
20 25 30

Gly Lys Thr Tyr Phe Phe Thr Trp Thr Asp Gln Ile Ser Arg Glu Ser
35 40 45

Arg Phe Leu Glu Arg Leu Ala Phe Ile Val Ser Glu Asn Cys Leu Ile
50 55 60

Phe Leu Ile His Ala Ile Thr Gly Gln
65 70

(2) INFORMATION FOR SEQ ID NO: 199:

15

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 289 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

Met Ser Gly Phe Ser Thr Glu Glu Arg Ala Ala Pro Phe Ser Leu Glu
1 5 10 15

Tyr Arg Val Phe Leu Lys Asn Glu Lys Gly Gln Tyr Ile Ser Pro Phe
20 25 30

His Asp Ile Pro Ile Tyr Ala Asp Lys Asp Val Phe His Met Val Val
35 40 45

Glu Val Pro Arg Trp Ser Asn Ala Lys Met Glu Ile Ala Thr Lys Asp
50 55 60

Pro Leu Asn Pro Ile Lys Gln Asp Val Lys Lys Gly Lys Leu Arg Tyr
65 70 75 80

Val Ala Asn Leu Phe Pro Tyr Lys Gly Tyr Ile Trp Asn Tyr Gly Ala
85 90 95

Ile Pro Gln Thr Trp Glu Asp Pro Gly His Asn Asp Lys His Thr Gly
100 105 110

Cys Cys Gly Asp Asn Asp Pro Ile Asp Val Cys Glu Ile Gly Ser Lys
115 120 125

Val Cys Ala Arg Gly Glu Ile Ile Gly Val Lys Val Leu Gly Ile Leu
130 135 140

Ala Met Ile Asp Glu Gly Glu Thr Asp Trp Lys Val Ile Ala Ile Asn
145 150 155 160

Val Asp Asp Pro Asp Ala Ala Asn Tyr Asn Asp Ile Asn Asp Val Lys
165 170 175

Arg Leu Lys Pro Gly Tyr Leu Glu Ala Thr Val Asp Trp Phe Arg Arg
180 185 190

Tyr Lys Val Pro Asp Gly Lys Pro Glu Asn Glu Phe Ala Phe Asn Ala
195 200 205

313

314

Glu Phe Lys Asp Lys Asp Phe Ala Ile Asp Ile Ile Lys Ser Thr His
210 215 220

5 Asp His Trp Lys Ala Leu Val Thr Lys Lys Thr Asn Gly Lys Gly Ile
225 230 235 240

Ser Cys Met Asn Thr Thr Leu Ser Glu Ser Pro Phe Lys Cys Asp Pro
245 250 255

10 Asp Ala Ala Arg Ala Ile Val Asp Ala Leu Pro Pro Cys Glu Ser
260 265 270

15 Ala Cys Thr Val Pro Thr Asp Val Asp Lys Trp Phe His His Gln Lys
275 280 285

Asn

20

(2) INFORMATION FOR SEQ ID NO: 200:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 625 amino acids

(B) TYPE: amino acid

(C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

30 Met Glu Ile Pro Gly Ser Leu Cys Lys Lys Val Lys Leu Ser Asn Asn
1 5 10 15

Ala Gln Asn Trp Gly Met Gln Arg Ala Thr Asn Val Thr Tyr Gln Ala
20 25 30

35 His His Val Ser Arg Asn Lys Arg Gly Gln Val Val Gly Thr Arg Gly
35 40 45

40 Gly Phe Arg Gly Cys Thr Val Trp Leu Thr Gly Leu Ser Gly Ala Gly
50 55 60

Lys Thr Thr Val Ser Met Ala Leu Glu Glu Tyr Leu Val Cys His Gly
65 70 75 80

45 Ile Pro Cys Tyr Thr Leu Asp Gly Asp Asn Ile Arg Gln Gly Leu Asn
85 90 95

Lys Asn Leu Gly Phe Ser Pro Glu Asp Arg Glu Glu Asn Val Arg Arg
100 105 110

50 Ile Ala Glu Val Ala Lys Leu Phe Ala Asp Ala Gly Leu Val Cys Ile
115 120 125

55 Thr Ser Phe Ile Ser Pro Tyr Thr Gln Asp Arg Asn Ala Arg Gln
130 135 140

Ile His Glu Gly Ala Ser Leu Pro Phe Phe Glu Val Phe Val Asp Ala
145 150 155 160

60 Pro Leu His Val Cys Glu Gln Arg Asp Val Lys Gly Leu Tyr Lys Lys

165 170 175

Ala Arg Ala Gly Glu Ile Lys Gly Phe Thr Gly Ile Asp Ser Glu Tyr
180 185 190

5 Glu Lys Pro Glu Ala Pro Glu Leu Val Leu Lys Thr Asp Ser Cys Asp
195 200 205

10 Val Asn Asp Cys Val Gln Gln Val Val Glu Leu Leu Gln Glu Arg Asp
210 215 220

Ile Val Pro Val Asp Ala Ser Tyr Glu Val Lys Glu Leu Tyr Val Pro
225 230 235 240

15 Glu Asn Lys Leu His Leu Ala Lys Thr Asp Ala Glu Thr Leu Pro Ala
245 250 255

Leu Lys Ile Asn Lys Val Asp Met Gln Trp Val Gln Val Leu Ala Glu
260 265 270

20 Gly Trp Ala Thr Pro Leu Asn Gly Phe Met Arg Glu Arg Glu Tyr Leu
275 280 285

25 Gln Cys Leu His Phe Asp Cys Leu Leu Asp Gly Gly Val Ile Asn Leu
290 295 300

Ser Val Pro Ile Val Leu Thr Ala Thr His Glu Asp Lys Glu Arg Leu
305 310 315 320

30 Asp Gly Cys Thr Ala Phe Ala Leu Met Tyr Glu Gly Arg Val Ala
325 330 335

Ile Leu Arg Asn Pro Glu Phe Phe Glu His Arg Lys Glu Glu Arg Cys
340 345 350

35 Ala Arg Gln Trp Gly Thr Cys Lys Asn His Pro Tyr Ile Lys Met
355 360 365

40 Val Met Glu Gln Gly Asp Trp Leu Ile Gly Gly Asp Leu Gln Val Leu
370 375 380

Asp Arg Val Tyr Trp Asn Asp Gly Leu Asp Gln Tyr Arg Leu Thr Pro
385 390 395 400

45 Thr Glu Leu Lys Gln Lys Phe Lys Asp Met Asn Ala Asp Ala Val Phe
405 410 415

Ala Phe Gln Leu Arg Asn Pro Val His Asn Gly His Ala Leu Leu Met
420 425 430

50 Gln Asp Thr His Lys Gln Leu Leu Glu Arg Gly Tyr Arg Arg Pro Val
435 440 445

55 Leu Leu Leu His Pro Leu Gly Gly Thr Lys Asp Asp Val Pro
450 455 460

Leu Met Trp Arg Met Lys Gln His Ala Val Leu Glu Gly Val
465 470 475 480

60 Leu Asn Pro Glu Thr Thr Val Val Ala Ile Phe Pro Ser Pro Met Met

315

485 490 495

Tyr Ala Gly Pro Thr Glu Val Gln Trp His Cys Arg Ala Arg Met Val
500 505 510

5 Ala Gly Ala Asn Phe Tyr Ile Val Gly Arg Asp Pro Ala Gly Met Pro
515 520 525

10 His Pro Glu Thr Gly Lys Asp Leu Tyr Glu Pro Ser His Gly Ala Lys
530 535 540

Val Leu Thr Met Ala Pro Gly Leu Ile Thr Leu Glu Ile Val Pro Phe
545 550 555 560

15 Arg Val Ala Ala Tyr Asn Lys Lys Lys Arg Met Asp Tyr Tyr Asp
565 570 575

20 Ser Glu His His Glu Asp Phe Glu Phe Ile Ser Gly Thr Arg Met Arg
580 585 590

Lys Leu Ala Arg Glu Gly Gln Lys Pro Pro Glu Gly Phe Met Ala Pro
595 600 605

25 Lys Ala Trp Thr Val Leu Thr Glu Tyr Tyr Lys Ser Leu Glu Lys Ala
610 615 620

Xaa
625

(2) INFORMATION FOR SEQ ID NO: 201:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 649 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

40 Met Ser Ala Ser Gln Asp Leu Glu Pro Lys Pro Leu Phe Pro Lys Pro
1 5 10 15

45 Ala Phe Gly Gln Lys Pro Pro Leu Ser Thr Glu Asn Ser His Glu Asp
20 25 30

Glu Ser Pro Met Lys Asn Val Ser Ser Ser Lys Gly Ser Pro Ala Pro
35 40 45

50 Leu Gly Val Arg Ser Lys Ser Gly Pro Leu Lys Pro Ala Arg Glu Asp
50 55 60

Ser Glu Asn Lys Asp His Ala Gly Glu Ile Ser Ser Leu Pro Phe Pro
65 70 75 80

55 Gly Val Val Leu Lys Pro Ala Ala Ser Arg Gly Gly Pro Gly Leu Ser
85 90 95

Lys Asn Gly Glu Glu Lys Lys Glu Asp Arg Lys Ile Asp Ala Ala Lys
100 105 110

316

Asn Thr Phe Gln Ser Lys Ile Asn Gln Glu Glu Leu Ala Ser Gly Thr
115 120 125

5 Pro Pro Ala Arg Phe Pro Lys Ala Pro Ser Lys Leu Thr Val Gly Gly
130 135 140

Pro Trp Gly Gln Ser Gln Glu Lys Lys Gly Asp Lys Asn Ser Ala
145 150 155 160

10 Thr Pro Lys Gln Lys Pro Leu Pro Pro Leu Phe Thr Leu Gly Pro Pro
165 170 175

Pro Pro Lys Pro Asn Arg Pro Pro Asn Val Asp Leu Thr Lys His
180 185 190

15 Lys Thr Ser Ser Gly Asn Ser Thr Ser Lys Gly Gln Thr Ser Tyr Ser
195 200 205

20 Thr Thr Ser Leu Pro Pro Pro Pro Ser His Pro Ala Ser Gln Pro
210 215 220

Pro Leu Pro Ala Ser His Pro Ser Gln Pro Pro Val Pro Ser Leu Pro
225 230 235 240

25 Pro Arg Asn Ile Lys Pro Pro Phe Asp Leu Lys Ser Pro Val Asn Glu
245 250 255

30 Asp Asn Gln Asp Gly Val Thr His Ser Asp Gly Ala Gly Asn Leu Asp
260 265 270

Glu Glu Gln Asp Ser Glu Gly Glu Thr Tyr Glu Asp Ile Glu Ala Ser
275 280 285

35 Lys Glu Arg Glu Lys Lys Arg Glu Lys Glu Lys Lys Arg Leu Glu
290 295 300

Leu Glu Lys Lys⁶ Glu Gln Lys Glu Lys Glu Lys Lys Glu Gln Ile
305 310 315 320

40 Lys Lys Lys Phe Lys Leu Thr Gly Pro Ile Gln Val Ile His Leu Ala
325 330 335

45 Lys Ala Cys Cys Asp Val Lys Lys Gly Gly Lys Asn Glu Leu Ser Phe Lys
340 345 350

Gln Gly Glu Gln Ile Glu Ile Ile Arg Ile Thr Asp Asn Pro Glu Gly
355 360 365

50 Lys Thr Leu Gly Arg Thr Ala Arg Gly Ser Tyr Gly Tyr Ile Lys Thr
370 375 380

Thr Ala Val Glu Ile Asp Tyr Asp Ser Leu Lys Leu Lys Lys Asp Ser
385 390 395 400

55 Leu Gly Ala Pro Ser Arg Pro Ile Glu Asp Asp Gln Glu Val Tyr Asp
405 410 415

Asp Val Ala Glu Gln Asp Asp Ile Ser Ser His Ser Gln Ser Gly Ser
420 425 430

317

Gly Gly Ile Phe Pro Pro Pro Asp Asp Ile Tyr Asp Gly Ile
435 440 445

Glu Glu Glu Asp Ala Asp Asp Gly Ser Thr Leu Gln Val Gln Glu Lys
450 455 460

Ser Asn Thr Trp Ser Trp Gly Ile Leu Lys Met Leu Lys Gly Lys Asp
465 470 475 480

Asp Arg Lys Lys Ser Ile Arg Glu Lys Pro Lys Val Ser Asp Ser Asp
485 490 495

Asn Asn Glu Gly Ser Ser Phe Pro Ala Pro Pro Lys Gln Leu Asp Met
500 505 510

Gly Asp Glu Val Tyr Asp Asp Val Asp Thr Ser Asp Phe Pro Val Ser
515 520 525

Ser Ala Glu Met Ser Gln Gly Thr Asn Val Gly Lys Ala Lys Thr Glu
530 535 540

Glu Lys Asp Leu Lys Lys Leu Lys Lys Gln Xaa Lys Xaa Lys Asp
545 550 555 560

Phe Arg Lys Lys Phe Lys Tyr Asp Gly Glu Ile Arg Val Leu Tyr Ser
565 570 575

Thr Lys Val Thr Thr Ser Ile Thr Ser Lys Lys Trp Gly Thr Arg Asp
580 585 590

Leu Gln Val Lys Pro Gly Glu Ser Leu Glu Val Ile Gln Thr Thr Asp
595 600 605

Asp Thr Lys Val Leu Cys Arg Asn Glu Glu Gly Lys Tyr Gly Tyr Val
610 615 620

Leu Arg Ser Tyr Leu Ala Asp Asn Asp Gly Glu Ile Tyr Asp Asp Ile
625 630 635 640

Ala Asp Gly Cys Ile Tyr Asp Asn Asp
645

45 (2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 55 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

Met Ala Trp Pro Ser Arg Ser Lys Met Phe Thr Leu Leu Pro Val Leu
1 5 10 15

Cys Tyr Leu Trp Ser Leu Trp Leu Pro Gln Phe Ser Trp Ile Gln Glu
20 25 30

Leu Lys Ala Val Leu Arg Asp Asp Gly Leu Ile Ser Ala Val Ala Trp
35 40 45

318

Asn Ala Glu Phe Gln Thr Cys
50 55

(2) INFORMATION FOR SEQ ID NO: 203:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 267 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

15 Met Val Lys Val Thr Phe Asn Ser Ala Leu Ala Gln Lys Glu Ala Lys
1 5 10 15

Lys Asp Glu Pro Lys Ser Gly Glu Glu Ala Leu Ile Ile Pro Pro Asp
20 25 30

Ala Val Ala Val Asp Cys Lys Asp Pro Asp Asp Val Val Pro Val Gly
35 40 45

Gln Arg Arg Ala Trp Cys Trp Cys Met Cys Phe Gly Leu Ala Phe Met
50 55 60

Leu Ala Gly Val Ile Leu Gly Gly Ala Tyr Leu Tyr Lys Tyr Phe Ala
65 70 75 80

Leu Gln Pro Asp Asp Val Tyr Tyr Cys Gly Ile Lys Tyr Ile Lys Asp
85 90 95

Asp Val Ile Leu Asn Glu Pro Ser Ala Asp Ala Pro Ala Ala Leu Tyr
100 105 110

Gln Thr Ile Glu Glu Asn Ile Lys Ile Phe Glu Glu Glu Val Glu
115 120 125

Phe Ile Ser Val Pro Val Pro Glu Phe Ala Asp Ser Asp Pro Ala Asn
130 135 140

Ile Val His Asp Phe Asn Lys Lys Leu Thr Ala Tyr Leu Asp Leu Asn
145 150 155 160

Leu Asp Lys Cys Tyr Val Ile Pro Leu Asn Thr Ser Ile Val Met Pro
165 170 175

Pro Arg Asn Leu Leu Glu Leu Ile Asn Ile Lys Ala Gly Thr Tyr
180 185 190

Leu Pro Gln Ser Tyr Leu Ile His Glu His Met Val Ile Thr Asp Arg
195 200 205

Ile Glu Asn Ile Asp His Leu Gly Phe Ile Tyr Arg Leu Cys His
210 215 220

Asp Lys Glu Thr Tyr Lys Leu Gln Arg Arg Glu Thr Ile Lys Gly Ile
225 230 235 240

Gln Lys Arg Glu Ala Ser Asn Cys Phe Ala Ile Arg His Phe Glu Asn
245 250 255 260

319

Lys Phe Ala Val Gln Thr Leu Ile Cys Ser Xaa
245 250 255 260 265

(2) INFORMATION FOR SEQ ID NO: 204:

10

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 315 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

15

Met Asp Leu Arg Gln Phe Leu Met Cys Leu Ser Leu Cys Thr Ala Phe
1 5 10 15

20

Ala Leu Ser Lys Pro Thr Gln Lys Lys Asp Arg Val His His Gln Pro
20 25 30

Gln Leu Ser Asp Lys Val His Asn Asp Ala Gln Ser Phe Asp Tyr Asp
35 40 45

25

His Asp Ala Phe Leu Gly Ala Gln Gln Ala Lys Thr Phe Asp Gln Leu
50 55 60

30

Thr Pro Gln Gln Ser Lys Gln Arg Leu Gly Lys Ile Val Ser Lys Ile
65 70 75 80

Asp Gly Asp Lys Asp Gly Phe Val Thr Val Asp Gln Leu Lys Asp Trp
85 90 95

35

Ile Lys Phe Ala Gln Lys Arg Trp Ile Tyr Gln Asp Val Gln Arg Gln
100 105 110

Trp Lys Gly His Asp Leu Asn Gln Asp Gly Leu Val Ser Trp Gln Gln
115 120 125

40

Tyr Lys Asn Ala Thr Tyr Gly Tyr Val Leu Asp Asp Pro Asp Pro Asp
130 135 140

45

Asp Gly Phe Asn Tyr Lys Gln Met Met Val Arg Asp Gln Arg Arg Phe
145 150 155 160

Lys Met Ala Asp Lys Asp Gly Asp Leu Ile Ala Thr Lys Gln Gln Phe
165 170 175

50

Thr Ala Phe Leu His Pro Gln Gln Tyr Asp Tyr Met Lys Asp Ile Val
180 185 190

Val Gln Gln Thr Met Gln Asp Ile Asp Lys Asn Ala Asp Gly Phe Ile
195 200 205

55

Asp Leu Gln Gln Tyr Ile Gly Asp Met Tyr Ser His Asp Gly Asn Thr
210 215 220

Asp Gln Pro Gln Trp Val Lys Thr Gln Arg Gln Phe Val Gln Phe
225 230 235 240

60

320

Arg Asp Lys Asn Arg Asp Gly Lys Met Asp Lys Gln Gln Thr Lys Asp
245 250 255

Trp Ile Leu Pro Ser Asp Tyr Asp His Ala Gln Ala Arg His
260 265 270

Leu Val Tyr Gln Ser Asp Gln Asn Lys Asp Gly Lys Leu Thr Lys Gln
275 280 285

Gln Ile Val Asp Lys Tyr Asp Leu Phe Val Gly Ser Gln Ala Thr Asp
290 295 300

Phe Gly Gln Ala Leu Val Arg His Asp Gln Phe
305 310 315

15

(2) INFORMATION FOR SEQ ID NO: 205:

20

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 207 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

25

Met Phe Asp Ala Val Leu Ile Leu Leu Leu Ile Pro Leu Lys Asp Lys
1 5 10

Leu Val Asp Pro Ile Leu Arg Arg His Gly Leu Leu Pro Ser Ser Leu
20 25 30

Lys Arg Ile Ala Val Gly Met Phe Phe Val Met Cys Ser Ala Phe Ala
35 40 45

Ala Gly Ile Leu Gln Ser Lys Arg Leu Asn Leu Val Lys Gln Lys Thr
50 55 60

Ile Asn Gln Thr Ile Gly Asn Val Val Tyr His Ala Ala Asp Leu Ser
65 70 75 80

Leu Trp Trp Gln Val Pro Gln Tyr Leu Leu Ile Gly Ile Ser Gln Ile
85 90 95

Phe Ala Ser Ile Ala Gly Leu Gln Phe Ala Tyr Ser Ala Ala Pro Lys
100 105 110

Ser Met Gln Ser Ala Ile Met Gly Leu Phe Phe Phe Ser Gly Val
115 120 125

Gly Ser Phe Val Gly Ser Gly Leu Leu Ala Leu Val Ser Ile Lys Ala
130 135 140

Ile Gly Trp Met Ser Ser His Thr Asp Phe Gly Asn Ile Asn Gly Cys
145 150 155 160

Tyr Leu Asn Tyr Tyr Phe Phe Leu Leu Ala Ala Ile Gln Gly Ala Thr
165 170 175

Leu Leu Leu Phe Leu Ile Ile Ser Val Lys Tyr Asp His His Arg Asp
180 185 190

60

321

322

His Gln Arg Ser Arg Ala Asn Gly Val Pro Thr Ser Arg Arg Ala
195 200 205

5

(2) INFORMATION FOR SEQ ID NO: 206:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 196 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

Met Arg Ser Arg Ile Arg Gln Phe Asp Ser Ser Thr Leu Asn Gln Ser
1 5 10 15

Val Arg Asn Thr Ile Met Arg Asp Leu Lys Ala Val Gly Lys Lys Phe
20 25 30

Met His Val Leu Tyr Pro Arg Lys Ser Asn Thr Leu Leu Arg Asp Trp
35 40 45

Asp Leu Trp Gly Pro Leu Ile Leu Cys Val Thr Leu Ala Leu Met Leu
50 55 60

Gln Arg Asp Ser Ala Asp Ser Gln Lys Asp Gly Gly Pro Gln Phe Ala
65 70 75 80

Glu Val Phe Val Ile Val Trp Phe Gly Ala Val Thr Ile Thr Leu Asn
85 90 95

Ser Lys Leu Leu Gly Gly Asn Ile Ser Phe Phe Gln Ser Leu Cys Val
100 105 110

Leu Gly Tyr Cys Ile Leu Pro Leu Thr Val Ala Met Leu Ile Cys Arg
115 120 125

Leu Val Leu Leu Ala Asp Pro Gly Pro Val Asn Phe Met Val Arg Leu
130 135 140

Phe Val Val Ile Val Met Phe Ala Trp Ser Ile Val Ala Ser Thr Ala
145 150 155 160

Phe Leu Ala Asp Ser Gln Pro Pro Asn Arg Arg Ala Leu Ala Val Tyr
165 170 175

Pro Val Phe Leu Phe Tyr Phe Val Ile Ser Trp Met Ile Leu Thr Phe
180 185 190

Thr Pro Gln Xaa
195

55

(2) INFORMATION FOR SEQ ID NO: 207:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 331 amino acids

(B) TYPE: amino acid

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

Met Ala Lys Asp Gln Ala Val Glu Asn Ile Leu Val Ser Pro Val Val
5 1 5 10 15

Val Ala Ser Ser Leu Gly Leu Val Ser Leu Gly Gly Lys Ala Thr Thr
20 25 30

Ala Ser Gln Ala Lys Ala Val Leu Ser Ala Glu Gln Leu Arg Asp Glu
35 40 45

Glu Val His Ala Gly Leu Gly Leu Leu Arg Ser Leu Ser Asn Ser
50 55 60

Thr Ala Arg Asn Val Thr Trp Lys Leu Gly Ser Arg Leu Tyr Gly Pro
65 70 75 80

Ser Ser Val Ser Phe Ala Asp Asp Phe Val Arg Ser Ser Lys Gln His
85 90 95

Tyr Asn Cys Glu His Ser Lys Ile Asn Phe Arg Asp Lys Arg Ser Ala
100 105 110

Leu Gln Ser Ile Asn Glu Trp Ala Ala Gln Thr Thr Asp Gly Lys Leu
115 120 125

Pro Glu Val Thr Lys Asp Val Glu Arg Thr Asp Gly Ala Leu Leu Val
130 135 140

Asn Ala Met Phe Phe Lys Pro His Trp Asp Glu Lys Phe His His Lys
145 150 155 160

Met Val Asp Asn Arg Gly Phe Met Val Thr Arg Ser Tyr Thr Val Gly
165 170 175

Val Met Met Met His Arg Thr Gly Leu Tyr Asn Tyr Tyr Asp Asp Glu
180 185 190

Lys Glu Lys Leu Gln Ile Val Glu Met Pro Leu Ala His Lys Leu Ser
195 200 205

Ser Leu Ile Ile Leu Met Pro His His Val Glu Pro Leu Glu Arg Leu
210 215 220

Glu Lys Leu Leu Thr Lys Glu Gln Leu Lys Ile Trp Met Gly Lys Met
225 230 235 240

Gln Lys Lys Ala Val Ala Ile Ser Leu Pro Lys Gly Val Val Glu Val
245 250 255

Thr His Asp Leu Gln Lys His Leu Ala Gly Leu Thr Glu Ala
260 265 270

Ile Asp Lys Asn Lys Ala Asp Leu Ser Arg Met Ser Gly Lys Lys Asp
275 280 285

Leu Tyr Leu Ala Ser Val Phe His Ala Thr Ala Phe Glu Leu Asp Thr
290 295 300

323

Asp Gly Asn Pro Leu Thr Arg Ile Thr Gly Gly Val Arg Thr Gln
305 310 315 320

Val Phe Tyr Ala Asp His Pro Phe Ile Ser Xaa
325 330

(2) INFORMATION FOR SEQ ID NO: 208:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 58 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

Met Cys Met Gln Leu Phe Gly Phe Leu Ala Phe Met Ile Phe Met Cys
1 5 10 15

Trp Val Gly Asp Val Tyr Pro Val Tyr Gln Pro Val Gly Pro Lys Gln
20 25 30

Tyr Pro Tyr Asn Asn Leu Tyr Leu Glu Arg Gly Gly Asp Pro Ser Lys
35 40 45

Glu Pro Glu Arg Val Val His Tyr Glu Ile
50 55

(2) INFORMATION FOR SEQ ID NO: 209:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 392 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

Met Asp Ala Leu Val Glu Asp Asp Ile Cys Ile Leu Asn His Glu Lys
1 5 10 15

Ala His Lys Arg Asp Thr Val Thr Pro Val Ser Ile Tyr Ser Gly Asp
20 25 30

Glu Ser Val Ala Ser His Phe Ala Leu Val Thr Ala Tyr Glu Asp Ile
35 40 45

Lys Lys Arg Leu Lys Asp Ser Glu Lys Glu Asn Ser Leu Leu Lys Lys
50 55 60

Arg Ile Arg Phe Leu Glu Glu Lys Leu Ile Ala Arg Phe Glu Glu Glu
65 70 75 80

Thr Ser Ser Val Gly Arg Glu Gln Val Asn Lys Ala Tyr His Ala Tyr
85 90 95

Arg Glu Val Cys Ile Asp Arg Asp Asn Leu Lys Ser Lys Leu Asp Lys
100 105 110

Met Asn Lys Asp Asn Ser Glu Ser Leu Lys Val Leu Asn Glu Gln Leu

324

115 120 125

Gln Ser Lys Glu Val Glu Leu Leu Gln Leu Arg Thr Thr Glu Val Glu Thr
130 135 140

Gln Gln Val Met Arg Asn Leu Asn Pro Pro Ser Ser Asn Trp Glu Val
145 150 155 160

Glu Lys Leu Ser Cys Asp Leu Lys Ile His Gly Leu Glu Gln Glu Leu
165 170 175

Glu Leu Met Arg Lys Glu Cys Ser Asp Lys Ile Glu Leu Gln Lys
180 185 190

Ala Lys Gln Thr Asp Pro Tyr Gln Glu Asp Asn Leu Lys Ser Arg Asp
195 200 205

Leu Gln Lys Leu Ser Ile Ser Ser Asp Asn Met Gln His Ala Tyr Trp
210 215 220

Glu Leu Lys Arg Glu Met Ser Asn Leu His Leu Val Thr Gln Val Gln
225 230 235 240

Ala Glu Leu Leu Arg Lys Leu Lys Thr Ser Thr Ala Ile Lys Lys Ala
245 250 255

Cys Ala Pro Val Gly Cys Ser Glu Asp Leu Gly Arg Asp Ser Thr Lys
260 265 270

Leu His Leu Met Asn Phe Thr Ala Thr Tyr Thr Arg His Pro Pro Leu
275 280 285

Leu Pro Asn Gly Lys Ala Leu Cys His Thr Thr Ser Ser Pro Leu Pro
290 295 300

Gly Asp Val Lys Val Leu Ser Glu Lys Ala Ile Leu Gln Ser Trp Thr
305 310 315 320

Asp Asn Glu Arg Ser Ile Pro Asn Asp Gly Thr Cys Phe Gln Glu His
325 330 335

Ser Ser Tyr Gly Arg Asn Ser Leu Glu Asp Asn Ser Trp Val Phe Pro
340 345 350

Ser Pro Pro Lys Ser Ser Glu Thr Ala Phe Gly Glu Thr Lys Thr Lys
355 360 365

Thr Leu Pro Leu Pro Asn Leu Pro Pro Leu His Tyr Leu Asp Gln His
370 375 380

Asn Gln Asn Cys Leu Tyr Lys Asn
385 390

(2) INFORMATION FOR SEQ ID NO: 210:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

60

325

(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

Met His His His Thr Gln Leu Met Phe Ile Tyr Leu Phe Ile Tyr Leu
1 5 10 15
Phe Ile Leu Gly Val Phe Phe Phe Phe Xaa
20 25

10

(2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

Met Asn Cys Ile Leu Leu Leu Tyr Leu Leu Ile Pro Thr Ile Ser Ile
1 5 10 15

Ser Val Val Pro Tyr Val Ala Leu Asn Ile Lys Tyr Ile Lys Glu Cys
20 25 30

25

Thr Glu Asn Ser Phe Tyr Xaa
35

30

(2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

Met Leu Leu His Leu Thr Ala Ala Phe Leu Gln Arg Ala Gln Phe Ser
1 5 10 15

Thr Tyr Phe Pro Gly Tyr Phe Asp Gly Gln Tyr Trp Leu Trp Trp Val
20 25 30

Phe Leu Val Leu Gly Phe Leu Leu Phe Leu Arg Gly Phe Ile Asn Tyr
35 40 45

Ala Lys Val Arg Lys Met Pro Glu Thr Phe Ser Asn Leu Pro Arg Thr
50 55 60

Arg Val Leu Phe Ile Tyr Xaa
65 70

50

55

(2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids
(B) TYPE: amino acid

60

326

(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

Met Leu Thr Phe Phe Met Ala Phe Leu Phe Asn Trp Ile Gly Phe Phe
1 5 10 15

Leu Ser Phe Cys Leu Thr Thr Ser Ala Ala Gly Arg Tyr Gly Ala Ile
20 25 30

Ser Gly Phe Gly Leu Ser Leu Ile Lys Trp Ile Leu Ile Val Arg Phe
35 40 45

Ser Thr Tyr Phe Pro Ala Phe Met Asn Ser Leu Ser Arg Ser Lys Arg
50 55 60

Thr Pro Ala Gly Ser Glu Ser Arg Cys Arg Thr Gln Arg Asn Asn His
65 70 75 80

Leu Leu Xaa

20

(2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 81 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

Met Ser Lys Arg Ser Ala Ser Phe Ile Leu Leu Pro Leu Leu Phe Leu
1 5 10 15

Lys Gly Ser Phe Ala Lys Leu Asn Ala Arg Ile Ser Asp Cys Leu Glu
20 25 30

Glu Arg Tyr Cys His Asn Leu Trp Met Val Phe Gln Gly Cys Val Ile
35 40 45

Thr Glu Leu His Leu Ser Arg Met Ser Lys Thr Leu Ser Ser Leu Cys
50 55 60

Tyr Asp Phe Val Ile Asn Val Tyr Ile Phe Phe Lys Phe Leu Asp Ile
65 70 75 80

Thr

50

(2) INFORMATION FOR SEQ ID NO: 215:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

Met Cys Ser Leu Phe Glu Ser Arg Phe Phe Cys Phe Val Leu Phe Ser

327

1 5 10 15
Glu Lys Ile Ile Gln Leu Cys Ala Ser Ile Ala Phe Leu Cys Phe Val
20 25 30
5 Lys His Val Pro Trp Pro Lys Trp Lys Arg Lys Cys Leu Ile Asn Ala
35 40 45
10 Phe
15 (2) INFORMATION FOR SEQ ID NO: 216:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 203 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 216:
Met Thr Leu Arg Pro Ser Leu Leu Pro Leu His Leu Leu Leu Leu Leu
1 5 10 15
25 Leu Leu Ser Ala Ala Val Cys Arg Ala Glu Ala Gly Leu Glu Thr Glu
20 25 30
Ser Pro Val Arg Thr Leu Gln Val Glu Thr Leu Val Glu Pro Pro Glu
35 40 45
30 Pro Cys Ala Glu Pro Ala Ala Phe Gly Asp Thr Leu His Ile His Tyr
50 55 60
Thr Gly Ser Leu Val Asp Gly Arg Ile Ile Asp Thr Ser Leu Thr Arg
65 70 75 80
Asp Pro Leu Val Ile Glu Leu Gly Gln Lys Gln Val Ile Pro Gly Leu
85 90 95
40 Glu Gln Ser Leu Leu Asp Met Cys Val Gly Glu Lys Arg Arg Ala Ile
100 105 110
Ile Pro Ser His Leu Ala Tyr Gly Lys Arg Gly Phe Pro Pro Ser Val
115 120 125
45 Pro Ala Asp Ala Val Val Gln Tyr Asp Val Glu Leu Ile Ala Leu Ile
130 135 140
Arg Ala Asn Tyr Trp Leu Lys Leu Val Lys Gly Ile Leu Pro Leu Val
145 150 155 160
Gly Met Ala Met Val Pro Pro Ser Trp Ala Ser Leu Gly Ile Thr Tyr
165 170 175
55 Thr Glu Arg Pro Ile Asp Pro Lys Ser Pro Lys Arg Ser Arg Lys
180 185 190
Arg Asn Glu Thr Arg Ala Lys Arg Asn Asn Lys
195 200
60 1

328

(2) INFORMATION FOR SEQ ID NO: 217:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 186 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 217:
Met Lys Thr Leu Met Thr Ile Cys Pro Gly Thr Val Leu Leu Val Phe
1 5 10 15
15 Ser Ile Ser Leu Trp Ile Ile Ala Ala Trp Thr Val Arg Val Cys Glu
20 25 30
Ser Pro Glu Ser Pro Ala Gln Pro Ser Gly Ser Ser Leu Pro Ala Trp
35 40 45
20 Tyr His Asp Gln Gln Asp Val Thr Ser Asn Phe Leu Gly Ala Met Trp
50 55 60
Leu Ile Ser Ile Thr Phe Leu Ser Ile Gly Tyr Gly Asp Met Val Pro
65 70 75 80
His Thr Tyr Cys Gly Lys Gly Val Cys Leu Leu Thr Gly Ile Met Gly
85 90 95
25 Ala Gly Cys Thr Ala Leu Val Val Ala Val Val Ala Arg Lys Leu Glu
100 105 110
Leu Thr Lys Ala Glu Lys His Val His Xaa Phe Met Met Asp Thr Gln
115 120 125
35 Leu Thr Lys Arg Ile Lys Asn Xaa Ala Ala Asn Val Leu Xaa Glu Thr
130 135 140
Trp Leu Ile Tyr Lys His Thr Lys Leu Leu Lys Ile Asp His Ala
145 150 155 160
40 Lys Val Arg Asn Thr Arg Gly Ser Ser Lys Tyr Pro Pro Val Glu
165 170 175
Glu Arg Gln Asp Gly Thr Glu Glu Ala Glu
180 185 190
45 (2) INFORMATION FOR SEQ ID NO: 218:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 218:
Met Lys Phe Leu Ala Val Leu Val Leu Leu Gly Val Ser Ile Phe Leu
1 5 10 15
60 Val Ser Ala Gln Asn Pro Thr Thr Ala Ala Pro Ala Asp Thr Tyr Pro

329

20 25 30
 Ala Thr Gly Pro Ala Asp Asp Glu Ala Pro Asp Ala Glu Thr Thr Ala
 35 40 45
 5 Ala Ala Thr Thr Ala Thr Thr Ala Pro Thr Thr Ala Thr Thr Ala
 50 55 60
 10 Ala Ser Thr Thr Ala Arg Lys Asp Ile Pro Val Leu Pro Lys Trp Val
 65 70 75 80
 Gly Asp Leu Pro Asn Gly Arg Val Cys Pro
 85 90

15

(2) INFORMATION FOR SEQ ID NO: 219:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 119 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

25 Met Ser Ser Ala Ala Ala Asp His Trp Ala Trp Leu Leu Val Leu Ser
 1 5 10 15
 Phe Val Phe Gly Cys Asn Val Leu Arg Ile Leu Leu Pro Ser Phe Ser
 20 25 30
 30 Ser Phe Met Ser Arg Val Leu Gln Lys Asp Ala Glu Gln Glu Ser Gln
 35 40 45
 Met Arg Ala Glu Ile Gln Asp Met Lys Gln Glu Leu Ser Thr Val Asn
 50 55 60
 Met Met Asp Glu Phe Ala Arg Tyr Ala Arg Leu Glu Arg Lys Ile Asn
 65 70 75 80
 40 Lys Met Thr Asp Lys Leu Lys Thr His Val Lys Ala Arg Thr Ala Gln
 85 90 95
 Leu Ala Lys Ile Lys Trp Val Ile Ser Val Val Ala Phe Tyr Val Leu Gln
 100 105 110
 45 Ala Ala Leu Met Ile Ser Leu Ile Trp Lys Tyr Tyr Ser Val Pro Val
 115 120 125
 50 Ala Val Val Pro Ser Lys Trp Ile Thr Leu Xaa
 130 135

(2) INFORMATION FOR SEQ ID NO: 220:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

330

Met Ser Ser Ala Ala Asp His Trp Ala Trp Leu Leu Val Leu Ser
 1 5 10 15
 5 Phe Val Phe Gly Cys Asn Val Leu Arg Ile Leu Leu Pro Ser Phe Ser
 20 25 30
 Ser Phe Met Ser Arg Val Leu Gln Lys Asp Ala Asp Arg Ser His Arg
 35 40 45
 10

15

(2) INFORMATION FOR SEQ ID NO: 221:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

25 Met Thr Ala Pro Leu Pro Leu Ser Gly Leu Ala Leu Phe Leu Ile
 1 5 10 15
 Val Phe Phe Ser Leu Gly Val Phe Cys Ile Cys His Ser His Trp Tyr
 20 25 30
 30 His Thr Leu Gln Gln Met Ala Gly Thr Glu Pro Lys Ala Leu Leu Leu
 35 40 45
 Ser Pro Pro Ala Ala Thr Thr Phe Val Thr Val Thr His Glu Val Trp
 50 55 60
 35 Lys Glu Gln Ala Leu Ala
 65 70

40

(2) INFORMATION FOR SEQ ID NO: 222:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 83 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

50 Met Thr Cys Ser Val Ala Leu Leu Leu Ile Leu Gly Leu Arg Cys Ser
 1 5 10 15
 Gly Val Arg Pro Gly Leu Val Gly Glu Gly His Asn Pro Ser Ser Leu Leu
 20 25 30
 55 Val Cys Leu Leu Leu Lys Asp Ser Arg Thr Asn Gln Gly Ser Cys Pro
 35 40 45
 Gly Gly Pro Trp Ser Glu Arg Asp Ile Glu Ser Val Thr Ser Asp Asn
 50 55 60
 60

331

Cys Glu Ala Thr Leu Gly Tyr Arg Asn His Ser Leu Pro Ser Asn Tyr
65 70 75 80
Tyr Asn Ser

(2) INFORMATION FOR SEQ ID NO: 223:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

Met Leu Thr Arg Ser Leu Lys Thr Leu Pro Ser Ala Cys Thr Ala Phe
1 5 10 15
Leu Leu Leu Phe Phe Leu Phe Ser Ser Gly Asp Pro Glu Leu Ser Cys
20 25 30

Ser Cys Thr Leu Arg Thr Gln Ser Ser Tyr Ser
35 40

(2) INFORMATION FOR SEQ ID NO: 224:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 184 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

Met Tyr Arg Pro Ser Val Leu Leu Leu Leu Leu Leu Arg His Gly
1 5 10 15

Ala Gln Gly Lys Pro Ser Pro Asp Ala Gly Pro His Gly Gln Gly Arg
20 25 30

Val His Gln Ala Ala Pro Leu Ser Asp Ala Pro His Asp Asp Ala His
35 40 45

Gly Asn Phe Gln Tyr Asp His Glu Ala Phe Leu Gly Arg Glu Val Ala
50 55 60

Lys Glu Phe Asp Gln Leu Thr Pro Glu Glu Ser Gln Ala Arg Leu Gly
65 70 75 80

Arg Ile Val Asp Arg Met Asp Arg Ala Gly Asp Gly Asp Gly Tyr Val
85 90 95

Ser Leu Ala Glu Leu Arg Ala Tyr Ile Ala His Thr Gln Gln Arg His
100 105 110

Ile Arg Asp Ser Val Ser Ala Ala Tyr Asp Thr Tyr Asp Thr Asp Arg
115 120 125

Asp Gly Arg Val Gly Tyr Glu Glu Leu Arg Asn Xaa Thr Tyr Gly His

332

130 135 140
Xaa Xaa Pro Xaa Glu Glu Phe His Asp Val Glu Asp Ala Glu Thr Tyr
145 150 155 160

Lys Lys Met Leu Xaa Arg Asp Glu Arg Arg Phe Arg Val Ala Asp Gln
165 170 175

Asp Gly Asp Ser Met Ala Thr Arg
180

(2) INFORMATION FOR SEQ ID NO: 225:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 71 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

Met Tyr Leu Phe Ile Leu Leu Ser Leu Ala Leu Ile Ser Asp Ala Met
1 5 10 15

Val Met Asp Glu Lys Val Lys Arg Ser Ser Leu Cys Tyr Thr Arg Leu Leu
20 25 30

Pro Ser Ala Thr Thr Met Pro Xaa Thr Arg Ile Thr Pro Asn Thr Gly
35 40 45

Ala Glu Xaa Ile Ser Val Xaa Thr Ala Thr Ser Ser Pro Ser Pro Leu
50 55 60

Thr Ala Pro Ile Met Tyr Pro
65 70

(2) INFORMATION FOR SEQ ID NO: 226:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

Met His Val Phe Val Leu Glu Ile Phe Leu
1 5 10

(2) INFORMATION FOR SEQ ID NO: 227:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 118 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

Met Ala Val Ala Thr Leu Ala Ser Glu Thr Leu Pro Leu Leu Ala Leu

333

1 5 10 15
 Thr Phe Ile Thr Asp Asn Ser Leu Val Ala Ala Gly His Asp Cys Phe
 20 25 30
 5 35 40 45
 Pro Val Leu Phe Thr Tyr Asp Ala Ala Ala Gly Met Leu Ser Phe Gly
 10 50 55 60
 Gly Arg Leu Asp Val Pro Lys Gln Ser Ser Gln Arg Gly Leu Thr Ala
 65 70 75 80
 Arg Glu Arg Phe Gln Asn Leu Asp Lys Ala Ser Ser Glu Gly Gly
 15 85 90 95
 Thr Ala Ala Gly Ala Gly Leu Asp Ser Leu His Lys Asn Ser Val Ser
 100 105 110
 Gln Ile Ser Val Leu Ser Gly Gly Lys Ala Lys Cys Ser Gln Phe Cys
 20 115 120 125
 Thr Thr Gly Met Asp Gly Gly Met Ser Ile Trp Asp Val Lys Ser Leu
 130 135
 25 140 145
 Glu Ser Ala Leu Lys Asp Leu Lys Ile Lys
 30 150 155
 (2) INFORMATION FOR SEQ ID NO: 228:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:
 35 1 5 10 15
 Leu Gly Ser Leu Ser Thr Ala Pro Ser Ser Ala Leu Pro Thr Leu Gly
 40 20 25
 Ala Arg Arg Thr Arg Ser Lys
 45 (2) INFORMATION FOR SEQ ID NO: 229:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 133 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:
 50 1 5 10 15
 Met Thr Tyr Phe Ser Gly Leu Leu Val Ile Leu Ala Phe Ala Ala Trp
 20 25 30
 55 35 40 45
 Val Ala Leu Ala Glu Gly Leu Gly Val Ala Val Tyr Ala Ala Ala Val
 60 45 50 55
 Leu Leu Gly Ala Gly Cys Ala Thr Ile Leu Val Thr Ser Leu Ala Met

334

Thr Ala Asp Leu Ile Gly Pro His Thr Asn Ser Gly Ala Phe Val Tyr
 50 55 60
 5 65 70 75 80
 Gly Ser Met Ser Phe Leu Asp Lys Val Ala Asn Gly Leu Ala Val Met
 85 90 95
 Ala Ile Gln Ser Leu His Pro Cys Pro Ser Glu Leu Cys Cys Arg Ala
 10 100 105 110
 Cys Val Ser Phe Tyr His Trp Ala Met Val Ala Val Thr Gly Gly Val
 115 120 125
 Gly Val Ala Ala Ala Leu Cys Leu Cys Ser Leu Leu Leu Trp Pro Thr
 130
 Arg Leu Arg Arg Xaa
 20 (2) INFORMATION FOR SEQ ID NO: 230:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:
 30 1 5 10 15
 Gly Lys Pro Thr Gly Lys Ser Leu Pro Leu Met Trp Met Ile Leu Met
 20 25
 Gln Pro Ile Ile Met Ile Ser Met Met Ser Asn Gly
 35 (2) INFORMATION FOR SEQ ID NO: 231:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 61 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:
 45 1 5 10 15
 Met Gln Gly Lys Phe Met Lys Val Gln Val Tyr Arg Phe Leu Lys Tyr
 20 25 30
 50 35 40 45
 Leu Leu Met Leu Leu Cys Met Phe Val Asn Arg Gly Met Ser Lys Asp
 45 50 55
 Ser Thr Lys Lys Pro Gly Gln Lys Leu Lys Val Ser Leu Gly Ser
 55 60
 Ile Leu Asn Met Lys Ser Gln Arg Pro Leu Ser Trp Cys
 60 (2) INFORMATION FOR SEQ ID NO: 232:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:
 65 1 5 10 15
 Leu Gly Ser Leu Ser Thr Ala Pro Ser Ser Ala Leu Pro Thr Leu Gly

335

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

Met Met Glu Arg Ser Met Met Ile Leu Leu Met Ala Ala Ser Met Thr
1 5 10 15

Met Thr Ser Thr Gln Leu Trp Ser Phe Cys Cys Val His
20 25

(2) INFORMATION FOR SEQ ID NO: 233:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

Met Trp Trp Gln Leu Ala Lys Lys Glu Glu Pro Gly Val Gly Ala Cys Ala
1 5 10 15

Leu Asp

(2) INFORMATION FOR SEQ ID NO: 234:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

Leu Xaa
1

(2) INFORMATION FOR SEQ ID NO: 235:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

Met Leu Ile Cys Arg Leu Val Leu Leu Ala Asp Pro Gly Pro Val Asn
1 5 10 15

Phe Met Val Arg Leu Phe Val Ile Val Met Phe Ala Trp Ser Ile
20 25 30

Val Ala Ser Thr Ala Phe Leu Ala Asp Ser Gln Pro Pro Asn Arg Arg
35 40 45 60

336

Ala Leu Ala Val Tyr Pro Val Phe Leu Phe Tyr Phe Val Ile Ser Trp
50 55 60

Met Ile Leu Thr Thr Phe Thr Pro Gln
65 70

(2) INFORMATION FOR SEQ ID NO: 236:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 96 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

Met Arg Ser Leu Leu Leu Ser Ala Phe Cys Leu Leu Glu Ala Ala
1 5 10 15

Leu Ala Ala Glu Val Lys Lys Pro Ala Ala Ala Ala Pro Gly Thr
20 25 30

Ala Glu Lys Leu Ser Pro Lys Ala Ala Thr Leu Ala Glu Arg Xaa Pro
35 40 45

Ala Trp Pro Ser Ala Cys Thr Arg Pro Trp Pro Arg Thr Arg Gln Trp
50 55 60

Arg Thr Ser Trp Cys His Pro Trp Trp Trp Pro Arg Arg Trp Gly Ser
65 70 75 80

Cys Arg Trp Ala Ala Arg Arg Pro Arg Arg Arg Arg Gln Cys
85 90 95

(2) INFORMATION FOR SEQ ID NO: 237:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 143 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

Met Arg Ser Leu Leu Leu Ser Ala Phe Cys Leu Leu Glu Ala Ala
1 5 10 15

Leu Ala Ala Glu Val Lys Lys Pro Ala Ala Ala Ala Ala Pro Gly Thr
20 25 30

Ala Glu Lys Leu Ser Pro Lys Ala Ala Thr Leu Ala Glu Arg Lys Arg
35 40 45 60

Pro Gly Leu Gln Leu Val Pro Gly His Gly Gln Gly Pro Gly Ser Gly
50 55 60

337

Glu His Pro Gly Val Thr Arg Gly Gly Gly Leu Val Ala Gly Ala Arg
65 70 75 80

Val Ala Gly Arg Gln Gly Asp His Gly Val Ala Gly Gln Gly Ser Ala
5 85 90 95

Glu Arg Arg Ala Ala Arg Arg Gly Gly Ala Arg Arg Pro Gly Arg
100 105 110

Ala Ala Ala Leu Thr Gln Gln Leu His Gly Ala Gln Arg Asp Leu Glu
115 120 125

Ala Gly Gln Pro Thr Val Arg Thr Gln Leu Ser Glu Leu Arg Xaa
130 135 140

15

(2) INFORMATION FOR SEQ ID NO: 238:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 142 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

Met Arg Ser Leu Leu Leu Ser Ala Phe Cys Leu Leu Glu Ala Ala
1 5 10 15

Leu Ala Ala Glu Val Lys Lys Pro Ala Ala Ala Ala Pro Gly Thr
20 25 30

Ala Glu Lys Leu Ser Pro Lys Ala Ala Thr Leu Ala Glu Arg Xaa Arg
35 40 45

Pro Gly Leu Gln Leu Val Pro Gly His Gly Gln Gly Pro Gly Ser Gly
50 55 60

Glu His Pro Gly Val Thr Arg Gly Gly Gly Leu Val Ala Gly Ala Arg
65 70 75 80

Val Ala Gly Arg Gln Gly Asp His Gly Val Ala Gly Gln Gly Ser Ala
85 90 95

Glu Arg Arg Ala Ala Ala Arg Arg Gly Gly Ala Arg Arg Pro Gly Arg
100 105 110

Ala Ala Ala Leu Thr Gln Gln Leu Xaa Gly Ala Gln Arg Asp Leu Glu
115 120 125

Ala Gly Gln Pro Thr Val Arg Thr Gln Leu Ser Glu Leu Arg
130 135 140

50

(2) INFORMATION FOR SEQ ID NO: 239:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

55

60

338

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

Asp Pro Glu Ala Ala Asp Ser Gly Glu Pro Gln Asn Lys Arg Thr Pro
1 5 10 15

Asp Leu Pro Glu Glu Glu Tyr Val Lys Glu Glu Ile Gln Glu Asn Glu
20 25 30

Glu Ala Val Lys Lys Met Leu Val Glu Ala Thr Arg Glu Phe Glu Glu
35 40 45

Val Val Val Asp Glu Ser
50

15

(2) INFORMATION FOR SEQ ID NO: 240:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

Gln Lys Leu Lys Arg Lys Ala Glu Glu Asp Pro Glu Ala Ala Asp Ser
1 5 10 15

Gly Glu Pro Gln Asn Lys Arg Thr Pro Asp Leu Pro Glu Glu Glu Tyr
20 25 30

Val Lys Glu Glu Ile Gln Glu Asn Glu Glu Ala Val Lys Lys Met Leu
35 40 45

Val Glu Ala Thr Arg Glu Phe Glu Glu Val Val Val Asp Glu Ser
50 55 60

35

(2) INFORMATION FOR SEQ ID NO: 241:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 113 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

Lys Ala Met Glu Lys Ser Ser Leu Thr Gln His Ser Trp Gln Ser Leu
1 5 10 15

Lys Asp Arg Tyr Leu Lys His Leu Arg Gly Gln Glu His Lys Tyr Leu
20 25 30

Leu Gly Asp Ala Pro Val Ser Pro Ser Ser Gln Lys Leu Lys Arg Lys
35 40 45

Ala Glu Glu Asp Pro Glu Ala Ala Asp Ser Gly Glu Pro Gln Asn Lys
50 55 60

Arg Thr Pro Asp Leu Pro Glu Glu Tyr Val Lys Glu Glu Ile Gln
65 70 75 80

55

60

339

Glu Asn Glu Glu Val Lys Lys Met Leu Val Glu Ala Thr Arg Glu
85 90 95
Phe Glu Glu Val Val Asp Glu Ser Pro Pro Asp Phe Glu Ile His
100 105 110
Ile

10 (2) INFORMATION FOR SEQ ID NO: 242:

15 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 148 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

20 Leu Pro Ser Tyr Asp Glu Ala Glu Arg Thr Lys Ala Glu Ala Thr Ile
1 5 10 15
Pro Leu Val Pro Gly Arg Asp Glu Asp Phe Val Gly Arg Asp Phe
20 25 30
Asp Asp Ala Asp Gln Leu Arg Ile Gly Asn Asp Gly Ile Phe Met Leu
35 40 45
Thr Phe Met Ala Phe Leu Phe Asn Trp Ile Gly Phe Leu Ser
50 55 60
Phe Cys Leu Thr Thr Ser Ala Ala Gly Arg Tyr Gly Ala Ile Ser Gly
65 70 75 80
Phe Gly Leu Ser Leu Ile Lys Trp Ile Leu Ile Val Arg Phe Ser Thr
85 90 95
Tyr Phe Pro Gly Tyr Phe Asp Gly Gln Tyr Trp Leu Trp Val Phe
100 105 110
Leu Val Leu Gly Phe Leu Leu Phe Leu Arg Gly Phe Ile Asn Tyr Ala
115 120 125
Lys Val Arg Lys Met Pro Glu Thr Phe Ser Asn Leu Pro Arg Thr Arg
130 135 140
Val Leu Phe Ile
145

50 (2) INFORMATION FOR SEQ ID NO: 243:

55 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

60

340

Ala Gly Arg Tyr Gly Ala Ile Ser Gly Phe Gly Leu Ser Leu Ile Lys
1 5 10 15
Trp Ile Leu Ile Val Arg Phe Ser
20

10 (2) INFORMATION FOR SEQ ID NO: 244:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

15 Met Lys His Leu Ser Ala Trp Asn Phe Thr Lys Leu Thr Phe Leu Gln
1 5 10 15
Leu Trp Glu Ile Phe Glu Gly Ser Val Glu Asn Cys Gln Thr Leu Thr
20 25 30
Ser Tyr Ser Lys Leu Gln Ile Lys Tyr Thr Phe Ser Arg Gly Ser Thr
35 40 45
Phe Tyr Ile
50

30 (2) INFORMATION FOR SEQ ID NO: 245:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 213 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

40 Phe Ser Ser Asp Phe Arg Thr Ser Pro Trp Glu Ser Arg Arg Val Glu
1 5 10 15
Ser Lys Ala Thr Ser Ala Arg Cys Gly Leu Trp Gly Ser Gly Pro Arg
20 25 30
Arg Arg Pro Ala Ser Gly Met Phe Arg Gly Leu Ser Ser Trp Leu Gly
35 40 45
Leu Gln Gln Pro Val Ala Gly Gly Gly Gln Pro Asn Gly Asp Ala Pro
50 55 60
Pro Glu Gln Pro Ser Glu Thr Val Ala Glu Ser Ala Glu Glu Glu Leu
65 70 75 80
Gln Gln Ala Gly Asp Gln Glu Leu Leu His Gln Ala Lys Asp Phe Gly
85 90 95
Asn Tyr Leu Phe Asn Phe Ala Ser Ala Ala Thr Lys Lys Ile Thr Glu
100 105 110
Ser Val Ala Glu Thr Ala Gln Thr Ile Lys Lys Ser Val Glu Gly
115

60

341

342

115 120 135
Lys Ile Asp Gly Ile Ile Asp Lys Thr Ile Ile Gly Asp Phe Gln Lys
130 135 140
5 Gln Gln Lys Lys Phe Val Glu Glu Gln His Thr Lys Lys Ser Glu Ala
145 150 155 160
Ala Val Pro Pro Trp Val Asp Thr Asn Asp Glu Glu Thr Ile Gln Gln
165 170 175
10 Gln Ile Leu Ala Leu Ser Ala Asp Lys Arg Asn Phe Leu Arg Asp Pro
180 185 190
15 Pro Ala Gly Val Gln Phe Asn Phe Asp Phe Asp Gln Met Tyr Pro Val
195 200 205
Ala Leu Val Met Leu
210
20
(2) INFORMATION FOR SEQ ID NO: 246:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:
Met Arg Phe Ala Leu Val Pro Lys Leu Val Lys Glu Glu Val Phe Trp
1 5 10 15
30 Arg Asn Tyr Phe Tyr Arg Val Ser Leu Ile Lys Gln Ser Ala Gln Leu
20 25 30
Thr Ala Leu Ala Ala Gln Gln Gln Ala Ala Gly Lys Gly Gly Glu Glu
35 40 45
40 Gln
(2) INFORMATION FOR SEQ ID NO: 247:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 76 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:
Ser Thr Ser Pro Gly Val Ser Glu Phe Val Ser Asp Ala Phe Asp Ala
1 5 10 15
55 Cys Asn Leu Asn Gln Glu Asp Leu Arg Lys Glu Met Glu Gln Leu Val
20 25 30
Leu Asp Lys Lys Gln Glu Thr Ala Val Leu Glu Glu Asp Ser Ala
35 40 45
60

Asp Trp Glu Lys Glu Leu Gln Gln Glu Leu Gln Tyr Glu Val Val
50 55 60
5 Thr Glu Ser Glu Lys Arg Asp Glu Asn Trp Asp Lys
65 70 75
(2) INFORMATION FOR SEQ ID NO: 248:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 62 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:
Ser Pro Trp Glu Ser Arg Arg Val Glu Ser Lys Ala Thr Ser Ala Arg
1 5 10 15
20 Cys Gly Leu Trp Gly Ser Gly Pro Arg Arg Arg Pro Ala Ser Gly Met
25 30
Phe Arg Gly Leu Ser Ser Trp Leu Gly Leu Gln Gln Pro Val Ala Gly
35 40 45
Gly Gly Gln Pro Asn Gly Asp Ala Pro Pro Glu Gln Pro Ser
50 55 60
(2) INFORMATION FOR SEQ ID NO: 249:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:
Pro Val Ala Gly Gly Gly Gln Pro Asn Gly Asp Ala Pro Pro Glu Gln
1 5 10 15
Pro Ser Glu Thr Val Ala Glu Ser Ala Glu Glu Glu Leu Gln Gln Ala
20 25 30
Gly Asp Gln Glu Leu Leu His Gln Ala Lys Asp Phe Gly Asn Tyr Leu
35 40 45
Phe Asn Phe Ala Ser Ala Ala Thr Lys Lys Ile Thr Glu Ser Val Ala
50 55 60
Glu
65
(2) INFORMATION FOR SEQ ID NO: 250:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 amino acids

343

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

5 Phe Gln Lys Glu Gln Lys Lys Phe Val Glu Gln Gln His Thr Lys Lys
1 5 10 15

Ser Glu Ala Ala Val Pro Pro Trp Val Asp Thr Asn Asp Glu Gln Thr
20 25 30

Ile Gln Gln Gln Ile Leu Ala Leu Ser Ala Asp Lys Arg Asn Phe Leu
35 40 45

Arg Asp Pro Pro Ala Gly Val Gln Phe Asn Phe Asp Phe Asp Gln Met
50 55 60

Tyr Pro Val Ala Leu Val Met Leu
65 70

(2) INFORMATION FOR SEQ ID NO: 251:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

30 Pro Phe Ile Cys Val Ala Arg Asn Pro Val Ser Arg Asn Phe Ser Ser
1 5 10 15

Pro Ile Leu Ala Arg Lys Leu Cys Glu Gly Ala Ala
20 25

(2) INFORMATION FOR SEQ ID NO: 252:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

45 Lys Glu Asp Pro Ala Asn Thr Val Tyr Ser Thr Val Glu Ile Pro Lys
1 5 10 15

Lys Met Glu Asn Pro His Ser Leu Leu Thr Met Pro Asp Thr Pro Arg
20 25 30

Leu

55

(2) INFORMATION FOR SEQ ID NO: 253:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 227 amino acids

60

344

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

5 Ala Ser Ala Val Leu Leu Asp Leu Leu Pro Asn Ser Gly Gly Glu Ala Gln
1 5 10 15

Ala Lys Lys Leu Gly Asn Asn Cys Val Phe Ala Pro Ala Asp Val Thr
20 25 30

Ser Glu Lys Asp Val Gln Thr Ala Leu Ala Leu Lys Gly Lys Phe
35 40 45

Gly Arg Val Asp Val Ala Val Asn Cys Ala Gly Ile Ala Val Ala Ser
50 55 60

Lys Thr Tyr Asn Leu Lys Lys Gly Gln Thr His Thr Leu Glu Asp Phe
65 70 75 80

Gln Arg Val Leu Asp Val Asn Leu Met Gly Thr Phe Asn Val Ile Arg
85 90 95

Leu Val Ala Gly Glu Met Gly Gln Asn Glu Pro Asp Gln Gly Gly Gln
100 105 110

Arg Gly Val Ile Ile Asn Thr Ala Ser Val Ala Ala Phe Glu Gly Gln
115 120 125

Val Gly Gln Ala Ala Tyr Ser Ala Ser Lys Gly Gly Ile Val Gly Met
130 135 140

Thr Leu Pro Ile Ala Arg Asp Leu Ala Pro Ile Gly Ile Arg Val Met
145 150 155 160

Thr Ile Ala Pro Gly Leu Phe Gly Thr Pro Leu Leu Thr Ser Leu Pro
165 170 175

Glu Lys Val Cys Asn Phe Leu Ala Ser Gln Val Pro Phe Pro Ser Arg
180 185 190

Leu Gly Asp Pro Ala Glu Tyr Ala His Leu Val Gln Ala Ile Ile Glu
195 200 205

Asn Pro Phe Leu Asn Gly Glu Val Ile Arg Leu Asp Gly Ala Ile Arg
210 215 220

Met Gln Pro
225

(2) INFORMATION FOR SEQ ID NO: 254:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

60 Ser Val Ala Ala Phe Glu Gly Gln Val Val Gly Gln Ala Ala Tyr Ser Ala

345

1 5 10 15
 Ser Lys Gly Ile Val Gly Met Thr Leu Pro Ile Ala 25
 5
 (2) INFORMATION FOR SEQ ID NO: 255:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 61 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:
 10 Ala Arg Arg Ser Gly Ala Glu Leu Ala Trp Asp Tyr Leu Cys Arg Trp 15
 1 5 10
 20 Ala Gln Lys His Lys Asn Trp Arg Phe Gln Lys Thr Arg Gln Thr Trp 30
 25 20
 Leu Leu Leu His Met Tyr Asp Ser Asp Lys Val Pro Asp Glu His Phe 45
 35
 25 Ser Thr Leu Leu Ala Tyr Leu Glu Gly Leu Gln Gly Arg 60
 50 55
 (2) INFORMATION FOR SEQ ID NO: 256:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:
 35 His Pro Ile Glu Trp Ala Ile Asn Ala Ala Thr Leu Ser Gln Phe Tyr 15
 1 5 10
 40 Ile Asn Lys Leu Cys Phe 20
 45 (2) INFORMATION FOR SEQ ID NO: 257:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:
 50 Cys Trp Ile Lys Tyr Cys Leu Thr Leu Met Gln Asn Ala Gln Leu Ser 15
 1 5 10
 55 Met Gln Asp Asn Ile Gly 20
 60

346

(2) INFORMATION FOR SEQ ID NO: 258:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:
 10 Lys Val Ser Tyr Leu Arg Pro Leu Asp Phe Glu Glu Ala Arg Glu Leu 15
 1 5 10
 15 Phe Leu Leu Gly Gln His Tyr Val Phe 25
 20
 (2) INFORMATION FOR SEQ ID NO: 259:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:
 20 Met Glu Arg Arg Cys Lys Met His Lys Arg Xaa Ile Ala Met Leu Glu 15
 1 5 10
 25 Pro Leu Thr Val Asp Leu Asn Pro Gln 25
 30
 (2) INFORMATION FOR SEQ ID NO: 260:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:
 35 Ser His Ile Val Lys Lys Ile Asn Asn Leu Asn Lys Ser Ala Leu Lys 15
 1 5 10
 40 Tyr Tyr Gln Leu Phe Leu Asp 20
 45
 (2) INFORMATION FOR SEQ ID NO: 261:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 64 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:
 50 Phe Thr His Leu Ser Thr Cys Leu Ser Leu Leu Leu Val Arg Met 15
 1 5 10
 55 Ser Gly Phe Leu Leu Ala Arg Ala Ser Pro Ser Ile Cys Ala Leu 15
 60

347

20 25 30
Asp Ser Ser Cys Phe Val Gln Glu Tyr Cys Ser Ser Tyr Ser Ser Ser
35 40 45
Cys Phe Leu His Gln His Phe Pro Ser Ser Leu Leu Asp His Leu Cys Gln
50 55 60

10

(2) INFORMATION FOR SEQ ID NO: 262:

15

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

Phe Leu Leu Leu Ala Arg Ala Ser Pro Ser Ile Cys Ala Leu Asp Ser
1 5 10 15

25

Ser Cys Phe Val Gln Glu Tyr
20

(2) INFORMATION FOR SEQ ID NO: 263:

30

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

Pro Asp Gly Arg Val Thr Asn Ile Pro Gln Gly Met Val Thr Asp Gln
1 5 10 15

40

Phe Gly Met Ile Gly Leu Leu Thr Phe Ile Arg Ala Ala Glu Thr Asp
20 25 30

Pro Gly Met Val His Leu Ala Leu Gly Ser Asp Leu Thr Thr Leu Gly
35 40 45

Leu Asn Leu Asn Ser
50

50

(2) INFORMATION FOR SEQ ID NO: 264:

55

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

Glu Asp Leu Leu Phe Tyr Leu Tyr Tyr Met Asn Gly Gly Asp Val Leu
1

348

1 5 10 15
Gln Leu Leu Ala Ala Val Glu Leu Phe Asn Arg Asp Tyr Arg Tyr His
20 25 30
Lys Glu Glu Arg Val Tyr Ile Thr Arg
35 40

10

(2) INFORMATION FOR SEQ ID NO: 265:

15

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

Val His Leu Ala Leu Gly Ser Asp Leu Thr Thr Leu Gly Leu Asn Leu
1 5 10 15

Asn Ser Pro Glu Asn Leu Tyr Pro
20

25

(2) INFORMATION FOR SEQ ID NO: 266:

30

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

His Asn Glu Asp Phe Pro Ala Leu Pro Gly Ser
1 5 10

40

(2) INFORMATION FOR SEQ ID NO: 267:

45

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 75 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

Gly Arg Ile Ile Asp Thr Ser Leu Thr Arg Asp Pro Leu Val Ile Glu
1 5 10 15

50

Leu Gly Gln Lys Gln Val Ile Pro Gly Leu Glu Gln Ser Leu Leu Asp
20 25 30

Met Cys Val Gly Glu Lys Arg Arg Ala Ile Ile Pro Ser His Leu Ala
35 40 45

Tyr Gly Lys Arg Gly Phe Pro Pro Ser Val Pro Ala Asp Ala Val Val
50 55 60

Gln Tyr Asp Val Glu Leu Ile Ala Leu Ile Arg

349

65

70

75

5 (2) INFORMATION FOR SEQ ID NO: 268:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

Ile His Tyr Thr Gly Ser Leu Val Asp Gly Arg Ile Ile Asp Thr Ser
1 5 10 15

15

20

(2) INFORMATION FOR SEQ ID NO: 269:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

Cys Glu Ser Pro Glu Ser Pro Ala Gln Pro Ser Gly Ser Ser Leu Pro
1 5 10 15Ala Trp Tyr His
20

35

(2) INFORMATION FOR SEQ ID NO: 270:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 95 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

Glu Glu Ala Gly Ala Gly Arg Cys Ser His Gly Gly Ala Arg Pro
1 5 10 15Ala Gly Leu Gly Asn Glu Gly Leu Gly Leu Gly Asp Pro Asp His
20 25 30Thr Asp Thr Gly Ser Arg Ser Lys Gln Arg Ile Asn Asn Trp Lys Glu
35 40 45Ser Lys His Lys Val Ile Met Ala Ser Ala Ser Ala Arg Gly Asn Gln
50 55 60Asp Lys Asp Ala His Phe Pro Pro Ser Lys Gln Ser Leu Leu Phe
65 70 75 80

60 Cys Pro Lys Ser Lys Leu His Ile His Arg Ala Glu Ile Ser Lys

95

85

90

350

5 (2) INFORMATION FOR SEQ ID NO: 271:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

Ser Lys Gln Arg Ile Asn Asn Trp Lys Glu Ser Lys His Lys Val Ile
1 5 10 15Met Ala Ser Ala Ser Ala Arg
20

20

(2) INFORMATION FOR SEQ ID NO: 272:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

Leu Phe His Trp Ala Cys Leu Asn Glu Arg Ala Ala Gln Leu Pro Arg
1 5 10 15Asn Thr Ala Xaa Ala Gly Tyr Gln Cys Pro Ser Cys Asn Gly Pro Ser
20 25 30

35

40 (2) INFORMATION FOR SEQ ID NO: 273:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 185 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

Phe Tyr Ile Tyr Tyr Arg Pro Thr Asp Ser Asp Asn Asp Ser Asp Tyr
1 5 10 15Lys Lys Asp Met Val Glu Gly Asp Lys Tyr Trp His Ser Ile Ser His
20 25 30Leu Gln Pro Glu Thr Ser Tyr Asp Ile Lys Met Gln Cys Phe Asn Glu
35 40 45Gly Gly Glu Ser Glu Phe Ser Asn Val Met Ile Cys Glu Thr Lys Ala
50 55 60

60 Arg Lys Ser Ser Gly Gln Pro Gly Arg Leu Pro Pro Thr Thr Leu Ala

351

65 70 75 80
Pro Pro Gln Pro Pro Leu Pro Gln Thr Ile Gln Arg Pro Val Gly Thr
85 90 95
5 Gly Ala Met Val Ala Arg Ser Ser Asp Leu Pro Tyr Leu Ile Val Gly
100 105 110
10 Val Val Leu Gly Ser Ile Val Leu Ile Ile Val Thr Phe Ile Pro Phe
115 120 125
Cys Leu Trp Arg Ala Trp Ser Lys Gln Lys His Thr Thr Asp Leu Gly
130 135 140
15 Phe Pro Arg Ser Ala Leu Pro Pro Ser Cys Pro Tyr Thr Met Val Pro
145 150 155 160
Leu Gly Gly Leu Pro Gly His Gln Ala Val Asp Ser Pro Thr Ser Val
165 170 175
20 Ala Ser Val Asp Gly Pro Val Leu Met
180 185

25 (2) INFORMATION FOR SEQ ID NO: 274:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

35 Tyr Ile Tyr Tyr Arg Pro Thr Asp Ser Asp Asn Asp Ser Asp Tyr Lys
1 5 10 15
Lys Asp Met Val Gln Gly Asp Lys Tyr Trp His Ser Ile Ser His Leu
20 25 30
40 Gln Pro Gln Thr Ser Tyr Asp Ile Lys Met Gln Cys Phe Asn Gln Gly
35 40 45
Gly Gln Ser Gln Phe Ser Asn Val Met Ile Cys Gln Thr Lys Ala Arg
50 55 60
45 Lys Ser
65

50 (2) INFORMATION FOR SEQ ID NO: 275:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

60 Asn Val Arg Ala Leu Leu His Arg Met Pro Gln Pro Pro Lys Ile Asn
1 5 10 15

352

Thr Ala Lys Phe Asn Asn Asn Lys Arg Lys Asn Leu Ser Leu
20 25 30

5 (2) INFORMATION FOR SEQ ID NO: 276:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 185 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

15 Asn Thr Asn Gln Arg Gln Ala Leu Gln Tyr Ala Lys Asn Phe Gln Pro
1 5 10 15
Phe Ala Leu Asn His Gln Lys Asp Ile Gln Val Leu Met Gly Ser Leu
20 25 30
20 Val Tyr Leu Arg Gln Gly Ile Gln Asn Ser Pro Tyr Val His Leu Leu
35 40 45
25 Asp Ala Asn Gln Trp Ala Asp Ile Cys Asp Ile Phe Thr Arg Asp Ala
50 55 60
Cys Ala Leu Leu Gly Leu Ser Val Gln Ser Pro Leu Ser Val Ser Phe
65 70 75 80
30 Ser Ala Gly Cys Val Ala Leu Pro Ala Leu Ile Asn Ile Lys Ala Val
85 90 95
Ile Gln Gln Arg Gln Cys Thr Gly Val Trp Asn Gln Lys Asp Gln Leu
100 105 110
35 Pro Ile Gln Val Asp Leu Gly Lys Lys Cys Trp Tyr His Ser Ile Phe
115 120 125
40 Ala Cys Pro Ile Leu Leu Arg Gln Gln Thr Thr Asp Asn Asn Pro Pro Met
130 135 140
Lys Leu Val Cys Gly His Ile Ile Ser Arg Asp Ala Leu Asn Lys Met
145 150 155 160
45 Phe Asn Gly Ser Lys Leu Lys Cys Pro Tyr Cys Pro Met Gln Gln Ser
165 170 175
Pro Gly Asp Ala Lys Gln Ile Phe Phe
180 185

50 (2) INFORMATION FOR SEQ ID NO: 277:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

60

353

Ser Tyr Leu Ser Ala Cys Phe Ala Gly Cys Asn Ser Thr Asn Leu Thr
1 5 10 15

Gly Cys Ala Cys Leu Thr Thr Val Pro Ala Glu Asn Ala Thr Val Val
5 20 25 30

Pro Gly Lys Cys Pro Ser Pro Gly Cys Gln Glu Ala Phe Leu Thr Phe
35 40 45

Leu Cys Val Met Cys Ile Cys Ser Leu Ile Gly Ala Met Ala Arg His
50 55 60

Pro
65

15

(2) INFORMATION FOR SEQ ID NO: 278:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

Pro Ser Val Ile Ile Leu Ile Arg Thr Val Ser Pro Glu Leu Lys Ser
1 5 10 15

Tyr Ala Leu Gly Val Leu Phe Leu Leu Leu Arg Leu Gly Phe Ile
20 25 30

Pro Pro Leu Ile Phe Gly Ala Gly Ile Asp Ser Thr Cys Leu Phe
35 40 45

Trp Ser Thr Phe Cys Gly Glu Gln Gly Ala Cys Val Leu Tyr Asp Asn
50 55 60

Val Val Tyr Arg Tyr Leu Tyr Val Ser Ile Ala Ile Ala Leu Lys Ser
65 70 75 80

Phe Ala Phe Ile

45

(2) INFORMATION FOR SEQ ID NO: 279:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 182 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

Gln Ser Leu Phe Thr Arg Phe Val Arg Val Gly Val Pro Thr Val Asp
1 5 10 15

Leu Asp Ala Gln Gly Arg Ala Arg Ala Ser Leu Cys Xaa Xaa Tyr Asn
20 25 30

Trp Arg Tyr Lys Asn Leu Gly Asn Leu Pro His Val Gln Leu Leu Pro

354

35 40 45

Glu Phe Ser Thr Ala Asn Ala Gly Leu Leu Tyr Asp Phe Gln Leu Ile
50 55 60

Asn Val Glu Asp Phe Gln Gly Val Gly Glu Ser Glu Pro Asn Pro Tyr
65 70 75 80

Phe Tyr Gln Asn Leu Gly Glu Ala Tyr Val Val Ala Leu Phe Met
85 90 95

Tyr Met Cys Leu Leu Gly Tyr Pro Ala Asp Lys Ile Ser Ile Leu Thr
100 105 110

Thr Tyr Asn Gly Gln Lys His Leu Ile Arg Asp Ile Ile Asn Arg Arg
115 120 125

Cys Gly Asn Asn Pro Leu Ile Gly Arg Pro Asn Lys Val Thr Val
130 135 140

Asp Arg Phe Gln Gly Gln Asn Asp Tyr Ile Leu Leu Ser Leu Val
145 150 155 160

Arg Thr Arg Ala Val Gly His Leu Arg Asp Val Arg Arg Leu Val
165 170 175

Ala Met Ser Arg Ala Arg
180

30

(2) INFORMATION FOR SEQ ID NO: 280:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

Leu Val Lys Glu Ala Lys Ile Ile Ala Met Thr Cys Thr His Ala Ala
1 5 10 15

Leu Lys Arg His Asp Leu Val Lys Leu Gly Phe Lys Tyr Asp Asn Ile
20 25 30

Leu Met Glu Glu Ala Ala Gln Ile Leu Glu Ile Glu Thr Phe Ile Pro
35 40 45

Leu Leu Leu Gln Asn Pro Gln Asp Gly Phe Ser Arg Leu Lys Arg Trp
50 55 60

Ile Met Ile Gly Asp His His Gln Leu Pro Pro Val Ile
65 70 75

55

(2) INFORMATION FOR SEQ ID NO: 281:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 125 amino acids

355

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

5 Asp Thr Tyr Pro Asn Glu Glu Lys Gln Gln Glu Arg Val Phe Pro Xaa
1 5 10 15
Xaa Ser Ala Met Val Asn Asn Gly Ser Leu Ser Tyr Asp His Glu Arg
20 25 30
Asp Gly Arg Pro Thr Glu Leu Gly Gly Cys Xaa Ala Ile Val Arg Asn
35 40 45
Leu His Tyr Asp Thr Phe Leu Val Ile Arg Tyr Val Lys Arg His Leu
50 55 60
Thr Ile Met Met Asp Ile Asp Gly Lys His Glu Trp Arg Asp Cys Ile
65 70 75
Glu Val Pro Gly Val Arg Leu Pro Arg Gly Tyr Tyr Phe Gly Thr Ser
85 90 95
Ser Ile Thr Gly Asp Leu Ser Asp Asn His Asp Val Ile Ser Leu Lys
100 105 110
Leu Phe Glu Leu Thr Val Glu Arg Thr Pro Glu Glu Glu
115 120 125

30 (2) INFORMATION FOR SEQ ID NO: 282:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 85 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

40 Leu Lys Arg Glu His Ser Leu Ser Lys Pro Tyr Gln Gly Val Gly Thr
1 5 10 15
Gly Ser Ser Ser Leu Trp Asn Leu Met Gly Asn Ala Met Val Met Thr
20 25 30
Gln Tyr Ile Arg Leu Thr Pro Asp Met Gln Ser Lys Gln Gly Ala Leu
35 40 45
Trp Asn Arg Val Pro Cys Phe Leu Arg Asp Trp Glu Leu Gln Val His
50 55 60
Phe Lys Ile His Gly Gln Gly Lys Lys Asn Leu His Gly Asp Gly Leu
65 70 75
Ala Ile Trp Tyr Thr
85

60 (2) INFORMATION FOR SEQ ID NO: 283:

356

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

5 Pro Gly Thr Leu Gln Cys Ser Ala Leu His His Asp Pro Gly Cys Ala
1 5 10 15
Asn Cys Ser Arg Phe Cys Arg Asp Cys Ser Pro Pro Ala Cys Gln Cys
20 25 30

15 (2) INFORMATION FOR SEQ ID NO: 284:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

25 Phe Leu Tyr Asp Val Leu Met Xaa His Glu Ala Val Met Arg Tyr His
1 5 10 15
Gln Ile Gln Leu Pro Asp Pro Glu Phe Pro Ser
20 25

30 (2) INFORMATION FOR SEQ ID NO: 285:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

40 Gly Trp Tyr Trp Cys Gly
1 5
(2) INFORMATION FOR SEQ ID NO: 286:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 129 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

55 Met Lys Val Gly Ala Arg Ile Arg Val Lys Met Ser Val Asn Lys Ala
1 5 10 15
His Pro Val Val Ser Thr His Trp Arg Trp Pro Ala Glu Trp Pro Gln
20 25 30

357

Met Phe Leu His Leu Ala Gln Glu Pro Arg Thr Glu Val Lys Ser Arg
35 40 45

5 Pro Leu Gly Leu Ala Gly Phe Ile Arg Gln Asp Ser Lys Thr Arg Lys
50 55 60

Pro Leu Glu Gln Glu Thr Ile Met Ser Ala Ala Asp Thr Ala Leu Trp
65 70 75 80

10 Pro Tyr Gly His Gly Asn Arg Glu His Gln Glu Asn Glu Leu Gln Lys
85 90 95

15 Tyr Leu Gln Tyr Lys Asp Met His Leu Leu Asp Ser Gly Gln Ser Leu
100 105 110

Gly His Thr His Thr Leu Gln Gly Ser His Asn Leu Thr Ala Leu Asn
115 120 125

Ile

20

(2) INFORMATION FOR SEQ ID NO: 287:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

Ser Leu His Lys Asn Ser Val Ser Gln Ile Ser Val Leu Ser Gly Gly
1 5 10 15

35 Lys Ala Lys Cys Ser Gln Phe Cys Thr Thr Gly Met Asp Gly Gly Met
20 25 30

Ser Ile Trp Asp Val Lys Ser Leu Glu Ser Ala Leu Lys Asp Leu Lys
35 40 45

Ile

40

(2) INFORMATION FOR SEQ ID NO: 288:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

55 Glu Ala Ser Lys Ser Ser His Ala Gly Leu Asp Leu Phe Ser Val Ala
1 5 10 15

Ala Cys His Arg Phe
20

60

358

(2) INFORMATION FOR SEQ ID NO: 289:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

10 Tyr Met Gly Lys Gly Ser Met Thr Gly Leu Ala Leu Lys His Met Phe
1 5 10 15

Glu Arg Ser Phe Thr
20

15

(2) INFORMATION FOR SEQ ID NO: 290:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

25 Val Thr Gly Ile Ile Asp Ser Leu Thr Thr Ile Ser Pro Lys Ala Ala Arg
1 5 10 15

Val Gly Leu Leu Gln Tyr Ser Thr Gln Val His
20 25

30

(2) INFORMATION FOR SEQ ID NO: 291:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

Thr Glu Phe Thr Leu Arg Asn Phe Asn Ser Ala Lys Asp Met Lys Lys
1 5 10 15

45 Ala Val Ala His Met Lys Tyr Met
20

50 (2) INFORMATION FOR SEQ ID NO: 292:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

Gly Lys Gly Ser Met Thr Gly Leu Ala Leu Lys His Met Phe Glu Arg
1 5 10 15

60

359

Ser Phe Thr Gln Gly Glu Gly Ala Arg Pro Phe
20 25

5 (2) INFORMATION FOR SEQ ID NO: 293:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

15 Ser Thr Arg Val Pro Arg Ala Ala Ile Val Phe Thr Asp Gly Arg Ala
1 5 10 15

Gln Asp Asp Val Ser Glu Trp Ala Ser Lys Ala Lys Ala Asn Gly Ile
20 25 30

20 Thr Met Tyr Ala Val Gly Val Gly Lys Ala Ile Glu
35 40

25 (2) INFORMATION FOR SEQ ID NO: 294:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

30 Gln Glu Leu Gln Glu Ile Ala Ser Glu Pro Thr Asn Lys His Leu Phe
1 5 10 15

35 Tyr Ala Glu Asp Phe Ser Thr Met Asp Glu Ile Ser Glu Lys Leu Lys
20 25 30

40 Lys Gly Ile Cys Glu Ala Leu Glu Asp Ser
35 40

45 (2) INFORMATION FOR SEQ ID NO: 295:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

50 Thr Gln Arg Leu Glu Glu Met Thr Gln Arg Met
1 5 10

55 (2) INFORMATION FOR SEQ ID NO: 296:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids

360

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

5 Pro Gln Gly Cys Pro Glu Gln Pro Leu His
1 5 10

10 (2) INFORMATION FOR SEQ ID NO: 297:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

15 Arg Cys Lys Lys Cys Thr Glu Gly Pro Ile Asp Leu Val Phe Val Ile
1 5 10 15

20 Asp Gly Ser Lys Ser Leu Gly Glu Glu Asn Phe Glu Val Val Lys Gln
20 25 30

25 Phe

30 (2) INFORMATION FOR SEQ ID NO: 298:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

35 Met Ala Ala Leu Leu Leu Arg His Val Gly Arg His Cys Leu Arg Ala
1 5 10 15

40 His Phe Ser Pro Gln Leu Cys Ile Arg Asn Ala Val Pro Leu Gly Thr
20 25 30

Thr Ala Lys Glu Glu Met Glu Arg Phe Thr Asn Lys Asn Ile Gly Ser
35 40 45

45 Asn Arg Pro Leu Ser Pro His Ile Thr Ile Tyr Ser
50 55 60

50 (2) INFORMATION FOR SEQ ID NO: 299:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

55 Val Phe Pro Leu Met Tyr His Thr Trp Asn Gly Ile Arg His Leu Met
1 5 10 15

361

362

Trp Asp Leu Gly Lys Gly Leu Lys Ile Pro Gln Leu Tyr Gln Ser Gly
20 25 30

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

Arg Val Trp Asp Val Arg Pro Phe Ala Pro Lys Glu Arg Cys Val Lys
1 5 10 15

10

Ile Phe Gln Gly Asn Val
20

15

(2) INFORMATION FOR SEQ ID NO: 304:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

His Asn Phe Glu Lys Asn Leu Leu Arg Cys Ser Trp Ser Pro Asp Gly
1 5 10 15

25

(2) INFORMATION FOR SEQ ID NO: 305:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

Val Lys Ser Leu Cys Leu Gly Pro Ala Leu Ile His Thr Ala Lys Phe
1 5 10 15

Ala Leu

40

(2) INFORMATION FOR SEQ ID NO: 306:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

Val Phe Pro Leu Met Tyr His Thr Trp Asn Gly Ile Arg His Leu Met
1 5 10 15

Trp Asp Leu Gly Lys Gly Leu
20

55

Val Arg Gly Arg Thr Val Leu Arg Pro Gly Leu Asp Ala Glu Pro Glu
1 5 10 15

Leu Ser Pro Glu
20

60

(2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:

363

(2) INFORMATION FOR SEQ ID NO: 307:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

10 Glu Gln Arg Val Leu Glu Arg Lys Leu Lys Lys Glu Arg Lys Lys Glu
1 5 10 15

Glu Arg Gln

15

(2) INFORMATION FOR SEQ ID NO: 308:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

25 Arg Leu Arg Glu Ala Gly Leu Val Ala Gln His Pro Pro
1 5 10

30 (2) INFORMATION FOR SEQ ID NO: 309:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

40 Gly Arg Ile Pro Ala Pro Ala Pro Ser Val Pro Ala Gly Pro Asp Ser
1 5 10 15

Arg

45 (2) INFORMATION FOR SEQ ID NO: 310:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

55 Thr Gly Cys Val Leu Val Leu Ser Arg Asn Phe Val Gln Tyr Ala Cys
1 5 10 15

60 Phe Gly Leu Phe Gly Ile Ile Ala Leu Gln Thr Ile Ala Tyr Ser Ile
20 25 30

364

Leu Tyr Asp Leu Lys Phe Leu Met Arg Asn
35 40

5 (2) INFORMATION FOR SEQ ID NO: 311:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 55 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

15 Ser Arg Ser Glu Gly Lys Ser Met Phe Ala Gly Val Pro Thr Met Arg
1 5 10 15

Glu Ser Ser Pro Lys Gln Tyr Met Gln Leu Gly Gly Arg Val Leu Leu
20 25 30

20 Val Leu Met Phe Met Thr Leu Leu His Phe Asp Ala Ser Phe Ser
35 40 45

Ile Val Gln Asn Ile Val Gly
50 55

25 (2) INFORMATION FOR SEQ ID NO: 312:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

35 Gly Thr Ala Glu Asp Phe Ala Asp Gln Phe Leu Arg Val Thr Lys Gln
1 5 10 15

40 Tyr Leu Pro His Val Ala Arg Leu Cys Leu Ile Ser Thr Phe Leu Glu
20 25 30

Asp Gly Ile Arg Met Thr Phe Gln Tyr Ser Glu Gln Arg Asp Tyr Ile
35 40 45

45 Asp Thr Thr Tyr Asn Cys Gly Tyr Leu Leu Ala Ser
50 55 60

50 (2) INFORMATION FOR SEQ ID NO: 313:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

55 Ala Ser Phe Leu Leu Ser Arg Thr Ser Thr Gly Thr Ala Leu Met Ile
1 5 10 15

60

Leu

5

(2) INFORMATION FOR SEQ ID NO: 314:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

Leu Met Arg Asn Glu Ser Arg Ser

1 5

10

15

20

(2) INFORMATION FOR SEQ ID NO: 315:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

Ala Ser Phe Leu Leu Ser Arg Thr Ser Trp Gly Thr Ala

1 5 10

30

(2) INFORMATION FOR SEQ ID NO: 316:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

Phe Ile Ser Phe Ala Asn Ser Arg Ser Ser Glu Asp Thr Lys Gln Met

1 5 10 15

Met Ser Ser Phe

20

40

45

(2) INFORMATION FOR SEQ ID NO: 317:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

Asp Pro Arg Arg Pro Asn Lys Val Leu Arg Tyr Lys Pro Pro Pro Ser

1 5 10 15

Glu Cys Asn Pro Ala Leu Asp Asp Pro Thr Pro

20 25

60

(2) INFORMATION FOR SEQ ID NO: 318:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

Asp Tyr Met Asn Leu Leu Gly Met Ile Phe Ser Met Cys Gly Leu Met

1 5 10 15

Leu Lys Leu Lys Trp Cys Ala Trp Val Ala Val Tyr Cys Ser

20 25 30

(2) INFORMATION FOR SEQ ID NO: 319:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

Met Leu Ser Ile Ser Ala Val Met Ser Tyr Leu Gln Asn Pro Gln

1 5 10 15

Pro Met Thr Pro Pro Trp

20

35

(2) INFORMATION FOR SEQ ID NO: 320:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

Ala Ala Gly Asp Gly Asp Val Lys Leu Gly Thr Leu Gly Ser Gly Ser

1 5 10 15

Glu Ser Ser Asn Asp Gly Gly Ser Glu Ser Pro Gly Asp Ala Gly Ala

20 25 30

Ala Ala Xaa Gly Gly Trp Ala Ala Ala Leu Ala Leu Leu Thr

35 40 45

Gly Gly Gly Glu

50

55

(2) INFORMATION FOR SEQ ID NO: 321:

(i) SEQUENCE CHARACTERISTICS:

367

(A) LENGTH: 177 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

5 Ala Ala Asp Asn Tyr Gly Ile Pro Arg Ala Cys Arg Asn Ser Ala Arg
1 5 10
10 Ser Tyr Gly Ala Ala Trp Leu Leu Xaa Pro Ala Gly Ser Ser Arg
20 25 30
15 Val Glu Pro Thr Gln Asp Ile Ser Ile Ser Asp Gln Leu Gly Gly Gln
35 40 45
20 Asp Val Pro Val Phe Arg Asn Leu Ser Leu Leu Val Val Gly Val Gly
50 55 60
25 Ala Val Phe Ser Leu Leu Phe His Leu Gly Thr Arg Glu Arg Arg Arg
65 70 75 80
30 Pro His Ala Xaa Glu Pro Gly Glu His Thr Pro Leu Leu Ala Pro Ala
85 90 95
35 Thr Ala Gln Pro Leu Leu Leu Trp Lys His Trp Leu Arg Arg Glu Xaa Ala
100 105 110
40 Phe Tyr Gln Val Gly Ile Leu Tyr Met Thr Thr Arg Leu Ile Val Asn
115 120 125
45 Leu Ser Gln Thr Tyr Met Ala Met Tyr Leu Thr Tyr Ser Leu His Leu
130 135 140
50 Pro Lys Lys Phe Ile Ala Thr Ile Pro Leu Val Met Tyr Leu Ser Gly
145 150 155 160
55 Phe Leu Ser Ser Phe Leu Met Lys Pro Ile Asn Lys Cys Ile Gly Arg
165 170 175
60 Asn
(2) INFORMATION FOR SEQ ID NO: 322:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 243 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 322:
5 Arg Ile Thr Asp Asn Pro Glu Gly Tyr Trp Leu Gly Arg Thr Ala Arg
1 5 10
10 Gly Ser Tyr Gly Tyr Ile Lys Thr Thr Ala Val Glu Ile Xaa Tyr Arg
20 25 30
15 Ser Leu Lys Leu Lys Lys Asp Ser Leu Gly Ala Pro Ser Arg Pro Ile
35 40 45
60 Glu Asp Asp Gln Glu Val Tyr Asp Asp Val Ala Glu Gln Asp Asp Ile

368

50 55 60

Ser Ser His Ser Gln Ser Gly Ser Gly Ile Phe Pro Pro Pro Pro
65 70 75 80

Asp Asp Asp Ile Tyr Asp Gly Ile Glu Glu Glu Asp Ala Asp Asp Gly
85 90 95

10 Phe Pro Ala Pro Pro Lys Gln Leu Asp Met Gly Asp Glu Val Tyr Asp
100 105 110

Asp Val Asp Thr Ser Asp Phe Pro Val Ser Ser Ala Glu Met Ser Gln
115 120 125

15 Gly Thr Asn Val Gly Lys Ala Lys Thr Glu Glu Lys Asp Leu Lys Lys
130 135 140

20 Leu Lys Lys Gln Xaa Lys Glu Xaa Lys Asp Phe Arg Lys Lys Phe Lys
145 150 155 160

Tyr Asp Gly Glu Ile Arg Val Leu Tyr Ser Thr Lys Val Thr Thr Ser
165 170 175

25 Ile Thr Ser Lys Lys Trp Gly Thr Arg Asp Leu Glu Val Lys Pro Gly
180 185 190

Glu Ser Leu Glu Val Ile Gln Thr Thr Asp Asp Thr Lys Val Leu Cys
195 200 205

30 Arg Asn Glu Glu Gly Lys Tyr Gly Tyr Val Leu Arg Ser Tyr Leu Ala
210 215 220

35 Asp Asn Asp Gly Glu Ile Tyr Asp Asp Ile Ala Asp Gly Cys Ile Tyr
225 230 235 240

Asp Asn Asp

40 (2) INFORMATION FOR SEQ ID NO: 323:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 106 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

45 Ser Met Ser Ala Leu Thr Arg Leu Ala Ser Phe Ala Arg Val Gly Gly
1 5 10 15

50 Arg Leu Phe Arg Ser Gly Cys Ala Arg Thr Ala Gly Asp Gly Val
20 25 30

55 Arg His Ala Gly Gly Gly Val His Ile Glu Pro Arg Tyr Arg Gln Phe
35 40 45

Pro Gln Leu Thr Arg Ser Gln Val Phe Gln Ser Glu Phe Phe Ser Gly
50 55 60

60 Leu Met Trp Phe Thr Ile Leu Trp Arg Phe Trp His Asp Ser Glu Glu

369

65 Val Leu Gly His Phe Pro Tyr Pro Asp Pro Ser Gln Trp Thr Asp Glu 80
85
Glu Leu Gly Ile Pro Pro Asp Asp Glu Asp 95
100 105

5


370

Applicant's or agent's file reference number	004PCT	International application	Unassigned
--	--------	---------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 73 . line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution	American Type Culture Collection
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit March 7, 1997	Accession Number 97923
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Isolation Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer		Authorized officer	

Applicant's or agent's file reference number	2004PCT	371	International application, Unassigned
			PCT No. 98/05311

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 73, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country)	
10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit	May 22, 1997
Accession Number	209071
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only

For International Bureau use only

<input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer	<input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer
--	--

Applicant's or agent's file reference number	2004PCT	372	International application, Unassigned
			PCT No. 98/05311

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 73, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country)	
10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit	February 25, 1998
Accession Number	209641
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only

For International Bureau use only

<input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer <i>Lawrence E. Ellis</i>	<input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer
--	--

373

Applicant's or agent's file reference number	Z004PCT	International application number	Unassigned PCT/US 98
--	---------	----------------------------------	-------------------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 75, line N/A	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
B. IDENTIFICATION OF DEPOSIT	
Name of depositary institution	American Type Culture Collection
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit	July 24, 1997
Accession Number	209179
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only	For International Bureau use only
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer V. V. V. V. V.	Authorized officer

374

Applicant's or agent's file reference number	Z004PCT	International application number	Unassigned PCT/US 98
--	---------	----------------------------------	-------------------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 77, line N/A	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
B. IDENTIFICATION OF DEPOSIT	
Name of depositary institution	American Type Culture Collection
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit	March 7, 1997
Accession Number	97924
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only	For International Bureau use only
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer V. V. V. V. V.	Authorized officer

Applicant's or agent's file reference number	Z004PCT	375	International application	Unassigned
--	---------	-----	---------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136i)

A. The indications made below relate to the microorganism referred to in the description on page 80 . line N/A	
B. IDENTIFICATION OF DEPOSIT	Further deposits are identified on an additional sheet <input type="checkbox"/>
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit March 13, 1997	Accession Number 97938
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

Applicant's or agent's file reference number	Z004PCT	376	International application	Unassigned
--	---------	-----	---------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136i)

A. The indications made below relate to the microorganism referred to in the description on page 80 . line N/A	
B. IDENTIFICATION OF DEPOSIT	Further deposits are identified on an additional sheet <input type="checkbox"/>
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit May 22, 1997	Accession Number 209072
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer <i>Thomas J. Allen</i>	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

377

Applicant's or agent's file reference number	Z004PCT	International application number	Unassigned
--	---------	----------------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 80 . line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit	September 4, 1997
Accession Number	209235
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only	For International Bureau use only
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer <i>Virginia L. Lley</i>	Authorized officer

378

Applicant's or agent's file reference number	Z004PCT	International application number	Unassigned
--	---------	----------------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 84 . line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit	August 28, 1997
Accession Number	209226
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only	For International Bureau use only
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer <i>Virginia L. Lley</i>	Authorized officer

Applicant's or agent's file reference number	Z004PCT	379	International application	Unassigned
			PCT/US 98	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136b)

A. The indications made below relate to the microorganism referred to in the description on page 84, line N/A	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country)	
10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit	March 13, 1997
Accession Number	97957
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
Authorized officer	Authorized officer
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:

Applicant's or agent's file reference number	Z004PCT	380	International application	Unassigned
			PCT/US 98	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136b)

A. The indications made below relate to the microorganism referred to in the description on page 84, line N/A	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country)	
10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit	May 22, 1997
Accession Number	209073
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
Authorized officer	Authorized officer
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:

381

382

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
 - (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;
 - (f) a polynucleotide which is a variant of SEQ ID NO:X;
 - (g) a polynucleotide which is an allelic variant of SEQ ID NO:X;
 - (h) a polynucleotide which encodes a species homologue of the SEQ ID NO:X;
 - (i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.
2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.
3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.
8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.
9. A recombinant host cell produced by the method of claim 8.
10. The recombinant host cell of claim 9 comprising vector sequences.
11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:
 - (a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
 - (b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;
 - (c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
 - (d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
 - (e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
 - (f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

383

- (g) a variant of SEQ ID NO:Y;
- (h) an allelic variant of SEQ ID NO:Y; or
- (i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

14. A recombinant host cell that expresses the isolated polypeptide of claim 11.

15. A method of making an isolated polypeptide comprising:
 (a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and
 (b) recovering said polypeptide.

16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:
 (a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and
 (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

- (a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and
- (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

384

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

- (a) contacting the polypeptide of claim 11 with a binding partner; and
- (b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

- (a) expressing SEQ ID NO:X in a cell;
- (b) isolating the supernatant;
- (c) detecting an activity in a biological assay; and
- (d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 22.

